Immigration, Innovation, and Growth*

Konrad B. Burchardi[†] **Thomas Chaney**[‡] **Tarek A. Hassan**[§]
Lisa Tarquinio[¶] **Stephen J. Terry**[‖]

March 2019

Abstract

Building upon endogenous growth theory, we show a causal impact of immigration on innovation and dynamism in US counties. In order to identify the causal impact of immigration, we use 130 years of detailed data on migrations from foreign countries to US counties to isolate quasi-random variation in the ancestry composition of US counties that results purely from the interaction of two forces: (i) changes over time in the relative attractiveness of different destinations within the US to the average migrant arriving at the time and (ii) the staggered timing of arrival of migrants from different origin countries. We then use this plausibly exogenous variation in ancestry composition to predict the total number of migrants flowing into each US county in recent decades. We show four main results. First, immigration has a positive impact on innovation, measured by patenting of local firms. Second, immigration has a positive impact on measures of local dynamism, as endogenous growth theory predicts. Third, the positive impact of immigration on innovation percolates over space, but spatial spillovers quickly die with distance. Fourth, the impact of immigration on innovation is stronger for more educated migrants.

JEL Classification: J61, O31, O40.

Keywords: migrations, innovation, patents, endogenous growth, dynamism.

---

*Preliminary and incomplete. We are grateful to Pascual Restrepo, Kevin Lang, Ray Fisman, Murat Celik, Daron Acemoglu, Klaus Desmet, and David Autor for helpful comments. We also thank seminar participants at the University of Toronto, Imperial College London, Southern Methodist University, and Boston University. Chaney is grateful for financial support from ERC grant N°337272–FiNet. All remaining mistakes are our own.

†Institute for International Economic Studies, 106 91 Stockholm, Sweden; E-mail: konrad.burchardi@iies.su.se.

‡Sciences Po, 28 rue des Saints Pères, 75007 Paris, France; E-mail: thomas.chaney@gmail.com.

§Boston University, NBER and CEPR, 270 Bay State Road, Boston MA 02215, USA; E-mail: thassan@bu.edu.

¶Boston University, 270 Bay State Road, Boston MA 02215, USA; E-mail: ltarq@bu.edu.

‖Boston University, 270 Bay State Road, Boston MA 02215, USA; E-mail: stephent@bu.edu.

1 Introduction

Does immigration cause more or less innovation and economic dynamism? In this paper, we answer this question in the context of international migrations to the US over the last three decades. We find a positive causal impact of immigration on both innovation and economic dynamism at the county level.

Models of endogenous growth suggest that immigration should increase both innovation and economic dynamism. As population grows with the inflow of migrants, more people can work towards more and possibly harder innovations (Jones, 1995; Bloom et al., 2017). As innovation is embedded in creative destruction at the microeconomic level, those new innovations facilitated by the inflow of migrants should also lead to more creative destruction and a higher level of overall economic dynamism (Aghion and Howitt, 1992; Grossman and Helpman, 1991; Klette and Kortum, 2004). If local returns to scale are large enough, the arrival of migrants should therefore cause innovation, economic dynamism, and economic growth within the receiving local communities.

Contrasting with these predictions of canonical growth theory, fierce political controversies surround the economic contribution of migrants: Are the new arrivals draining resources of their host communities and stifling innovation and economic dynamism?

A rigorous quantification of the causal impact of immigration on innovation and dynamism has proven elusive. The reason is that migrants do not allocate randomly across space, but instead likely tend to choose destinations that offer the best prospects for them and their families. In particular, it is plausible that migrants arriving in the US will tend to select into regions that are innovative, economically dynamic, and fast-growing, creating a spurious correlation between local immigration, local innovation, and local economic dynamism.

Our main contribution is to propose a formal identification strategy which allows us to quantify the causal impact of migrations on local innovation and dynamism. To do so, we use 130 years of detailed data on migrations from foreign countries to US counties. Our identification strategy combines a set of instruments for the pre-existing ethnic composition of US counties (Burchardi et al., forthcoming) with a canonical shift-share approach (Altonji and Card, 1991) to construct a valid instrument for migrations in the last 30 years. In a first step, we isolate plausibly exogenous variations in the number of residents of a US county with ancestry from different foreign countries, following Burchardi et al. (forthcoming). In a second step, we use these exogenous components of pre-existing ancestry shares to predict where recent migrants will settle within the US following Altonji and Card (1991). Doing so, we are immune to the potential critique that where migrants settle within the US, both in recent decades (the distribution of immigrants) and in the more distant past (the distribution of ancestry), may be correlated with unobserved factors that also affect measures of local innovation and dynamism.

In our first step, we use the interaction of time-series variation in the relative attractiveness of different destinations within the US with the staggered timing of arrival of migrants from different origins to isolate quasi-random variation in the settlement of migrants across US counties. Implicitly, we assume that migrations are in part driven by (i) a push factor, causing emigration from a given foreign country to the entire US, and (ii) a pull factor, causing immigration into a given US county from all origins. To further ensure that our predicted immigration measure is not contaminated by endogenous unobserved factors, we carefully leave out large population groups when predicting ancestry. In particular, as our focus is on immigration to the US after 1970, primarily originating from non-European countries, we use European migrations to predict where non-European migrants will settle. In other words, we predict that US counties that were attractive to migrants from Europe in a period when a large number of migrants from a given non-European origin country were arriving in the US will receive a large number of migrants from that origin country. Iterating this procedure over 130 years, we are able to isolate quasi-random variation in the distribution of ancestry across US counties in 1970.

In our second step, we use this predicted pre-existing distribution of ancestry to predict where new migrants arriving in the US after 1970 will settle. Implicitly, we assume that new migrants will tend to settle in locations where a large community from the same ethnic background already lives. So if a large community with ancestry from origin country o already resides in destination county d, and many migrants from o arrive in the US, we will predict a large inflow of migrants from o to d. Summing over all possible origin countries, we are therefore able to predict the total number of migrants flowing into different US counties. This predicted immigration is plausibly exogenous to any unobserved factor which may make a destination US county more innovative and dynamic after 1970.

This formal identification strategy allows us to reach four main conclusions.

First, we find a strong and significant causal impact of immigration on the number of patents filed per person: the arrival of 30,000 additional immigrants in a given county adds one patent per 1,000 inhabitants over a 5-year period. Put differently, a one standard deviation increase in the number of migrants causes a 21% increase relative to the sample mean in the number of patents filed per capita over 5 years.

Second, we also find a strong and significant causal impact of immigration on measures of economic dynamism and growth at the local level. For our measures of economic dynamism, or creative destruction, we use several variables, each shedding light on one aspect of economic dynamism: a one standard deviation increase in a county's number of immigrants results in a 2.1 percentage point increase in its job creation rate and a 1.8 percentage point increase in its job destruction rate relative the previous 5-year period. Consistent with this pattern, the skewness of the job creation rate is also positively associated with immigration, as endogenous growth theory predicts (Klette and Kortum, 2004). Finally, average local wages also rise significantly as a result of immigration, suggesting that immigration does not only affect innovation and creative destruction, but also the overall level of economic growth at the local level.

Third, we also find evidence that the positive effect of immigration on innovation and growth diffuses over space, but this spatial diffusion dies out quickly with distance. For instance, if more migrants settle in the US state surrounding county d, innovation in d increases significantly. This effect on immigration in surrounding counties is however substantially muted compared to the direct effect of immigration in county d: while a one standard deviation increase in the number of migrants at the state level has about the same effect on innovation as a one standard deviation increase in the number of migrants at the county level, the effect per migrant is almost three orders of magnitude smaller at the state than at the county level, simply because there are many more migrants at the state than at the county level.

Fourth, we find that the positive effect of immigration on innovation and growth is significantly stronger for more educated migrants. We are able to reach this conclusion because our identification strategy allows us to construct separate instruments for migrants from each origin at each point in time. This versatility is one of the strengths of our identification strategy. To separately identify the impact of the total number of incoming migrants from that of their education level, we rely on the fact that the level of education of migrants varies across countries of origin, and over time. We find large heterogeneity in the impact of immigration on innovation as we exogenously vary the education level of migrants. For instance, a county receiving relatively uneducated migrants (around the 3rd decile in the distribution of years of schooling among incoming migrants) would see almost no impact on innovation, while a county receiving highly educated migrants (around the 8th decile) would see one additional patent per 1,000 inhabitants over 5 years for every 2,500 immigrants.

Related Literature. Our paper builds upon several strands of the theoretical and applied literature. First, our empirical study is motivated by theories of endogenous growth. While the seminal contribution of Romer (1990) predicts that a larger population, in levels, ought to increase the rate of growth of the economy, subsequent refinements (Jones, 1995, 1999) in so-called semi endogenous growth theory predict a positive link between the growth rate of population and economic growth.[1] In these models, a larger population, or a higher population growth rate, allow an economy to grow because increasing returns to scale in the technology for innovating overcome decreasing returns to scale in production (Solow, 1956). Another key insight from this literature is that a larger population of (potential) innovators may be better able to achieve increasingly hard innovations. Bloom et al. (2017) show evidence that new ideas are in fact becoming harder to produce over time, so that an increasing amount of resources are needed to continue innovating. We contribute to this literature first by showing a positive impact of immigration –one channel through which population grows - on innovation and growth. We also show evidence that the scale effects necessary for endogenous growth in Jones (1995) operate and are statistically identifiable at the local level. Finally, we show that a key input in the technology to produce innovation is human capital, whereby more educated migrants contribute more to innovation than less educated migrants.

Endogenous growth theory also ties the innovation and growth process to creative destruction, as in the seminal contributions of Aghion and Howitt (1992), Grossman and Helpman (1991), and Klette and Kortum (2004). We confirm the predictions from those theories: an increase in population, induced by immigration, which feeds the innovation process, also increases measures of creative destruction. US counties receiving more migrants experience both higher rates of job creation and destruction, and they see more positively skewed growth, with more "superstar" or right-tail growth of industries in their area.

Second, we contribute to the recent but growing literature on the link between immigration and growth. Using data on the age of mass migration and the expansion of railways at the time, Sequeira et al. (forthcoming) show a long-run effect of immigration on local economic development. Akcigit et al. (2017) show that many inventors are immigrants. Bernstein et al. (2018) also show that many inventions originate in the work of immigrants. Peters (2017) investigates the link between historical refugees in Germany and industrialisation. Kerr and Kerr (2016) study the special case of immigrant entrepreneurship. Finally, Tabellini (2018) shows that while historical immigration to the US has fostered development, it has also sparked a political backlash.

Third, we contribute to the empirical literature which has investigated the decline in dy-

[1] This purposefully brief description does not do full justice to the rich debate on scale effects in growth, with other important contributions including Peretto (1998) and Young (1998).

namism, in particular in the US. Using aggregate trend data and a structural approach, Karahan et al. (2016) show the declining dynamism in the US may potentially be due to lower population growth. Using US census microdata, Decker et al. (2014) document the declining dynamism in the US. Gordon (2012) contemplates the idea that growth may have run out in the US. Alon et al. (2018) suggest that the decline in US dynamism may be due to older firms. Using data on aggregate trends, and guided by the theory, Hopenhayn et al. (2018) propose that the decline in dynamism in the US is due to older firms and a lower population growth rate. Hathaway and Litan (2014) also document declining measures of dynamism in the US. We contribute to this literature by showing causal evidence that population growth, and in particular population growth driven by immigration, may counteract the decline in dynamism in the US economy.

The remainder of this paper is structured as follows. Section 2 introduces our data. Section 3 lays out our strategy for identification and isolates empirically quasi-random variation in migrations to US counties. Section 4 formally estimates the causal effect of immigration on innovation, economic dynamism, and income growth. Section 5 tests for geographic spill-overs in the effect of immigration on innovation and disentangles the impact of high-skilled from that of low-skilled migration. Section 6 concludes.

2 Data

We collect detailed data on migrations, ancestry, the education level of migrants, patents issued, and measures of dynamism of local firms and local labor markets. Below is a description of our data sources, and the construction of our main variables. Further details on the construction and sources of the data are given in Appendix A.

Immigration and Ancestry. Following Burchardi et al. (forthcoming), our immigration and ancestry data are constructed from the individual files of the Integrated Public Use Microdata Series (IPUMS) samples of the 1880, 1900, 1910, 1920, 1930, 1970, 1980, 1990, and 2000 waves of the US census, and the 2006-2010 five-year sample of the American Community Survey. We weigh observations using the personal weights provided by these data sources. Appendix Table 1 summarizes specific samples and weights used. We cannot use data from the 1940, 1950 and 1960 censuses, because these did not collect information on the year of immigration. The original 1890 census files were lost in a fire.

Throughout the paper, we use $t-1$ and t to denote the end years of consecutive 5-year periods,[2] o for the foreign country of origin, and d for the US destination county. We construct the number of migrants from origin o to destination d at time t, $I^t_{o,d}$, as the number of respondents born in o who live in d in a given census year and emigrated to the United States between $t-1$ and t. The exception to this rule is the 1880 census (the first in our sample), which also did not record the year of immigration. The variable $I^{1880}_{o,d}$ instead measures the number of residents who were either born in o or whose parents were born in o, thus covering the two generations of immigrants arriving prior to 1880.[3] Since 1980, respondents have also been asked about their primary ancestry in both the US Census and the American Community Survey, with the option to provide multiple answers. $Ancestry^t_{o,d}$ corresponds to the number of individuals residing in d at time t who report o as first ancestry. Note that this measure captures self-reported (recalled) ancestry.[4]

The respondents' residence is recorded at the level of historic counties, and at the level of historic county groups or PUMAs from 1970 onwards. Whenever necessary we use contemporaneous population weights to transition data from the historic county group or PUMA level to the historic county, and then use area weights to transition data from the historic county level to the 1990 US county level. The respondents' stated ancestry (birthplace) often, but not always, directly corresponds to foreign countries in their 1990 borders (for example, "Spanish" or "Denmark"). When no direct mapping exists (for example, "Basque" or "Lapland"), we construct transition matrices that map data from the answer level to the 1990 foreign country level, using approximate population weights where possible and approximate area weights otherwise. In the few cases when answers are imprecisely specific or such a mapping cannot be constructed (for example, "European" or "born at sea"), we omit the data.[5] The resulting dyadic dataset covers 3,141 US counties, 195 foreign countries, and 10 census waves.

Innovation. We use patent data to measure innovation. Starting from the universe of patent microdata provided by the US Patent and Trademark Office (USPTO) from 1975 until 2010, we study corporate utility patents with US assignees, around 4.7 million observations. We convert assignee locations provided by the USPTO in coordinate form to 2010 US counties, tabulating the number of corporate utility patents granted to assignees in each county in each year of the

[2] Due to data limitations in the reporting of year of immigration in the 1970-1990 censuses, the length of periods vary slightly over this time frame.

[3] If the own birthplace is in the United States, imprecisely specific (e.g., a continent), or missing, we instead use the parents' birthplace, assigning equal weights to each parent's birthplace.

[4] See Duncan and Trejo (2017) for recent evidence on recalled versus factual ancestry in CPS data.

[5] Appendix Tables 2 and 3 report summary statistics on these data transitions, including the share of affected respondents. Appendix A.1 provides a detailed description of the data transformation.

sample, and then use area weights to transition to 1990 US counties. In the earlier periods, when there was more personalized US innovation, inventor location is the natural measure of geography (Akcigit et al., 2017); however, in recent years the overwhelming majority of patents are assigned to corporations, making assignees the natural location measure for our purposes. We sum patent flows over five-year periods, with the measure in t corresponding to the sum of patents in a given county d over the five years between $t-1$ and t. We then scale this measure by the 1970 population of county d from the Census microdata to yield a five-year patents per capita variable. The change in the flow of patents per capita from period $t-1$ to t is our primary outcome of interest.[6]

Dynamism. A growing empirical literature emphasizes that measures of dynamism and creative destruction in the United States have declined in recent decades (Decker et al., 2014). Our dynamism measures come from two sources, motivated by the prior work on this subject. The first dataset - the US Business Dynamism Statistics (BDS) database from the US Census - represents measures computed from the underlying Longitudinal Business Data microdata on the employment levels of the universe of US business establishments. The BDS data include job creation and job destruction rates (gross flows representing the ratio of the number of jobs created or destroyed as a fraction of employment) at the yearly level and spanning 1977–2015. The sum of these measures is known as the job reallocation rate, and the difference between them is the net employment growth rate. We apportion the native MSA geography to 1990 US counties by population. Our main dynamism outcomes of interest from the BDS data in county d in period t correspond to the change in either job creation or job destruction rates from $t-1$ to t.

In addition to measures of gross employment flows, the dynamism literature also emphasizes a decline over time in the skewness of employment growth rates, i.e., a decline in the relative importance of "superstar" growth performance in driving US employment dynamism. In this spirit, we construct growth rate skewness measures starting from the US Census County Business Patterns (CBP) dataset. The raw data contain county by year by 4-digit industry employment levels from 1985 to 2010. For each county and year, we compute the Kelley Skewness of employment growth rates across 4-digit sectors. This measure gives a sense of whether certain strongly performing industries drive overall employment growth in that period and lo-

[6] We manually check the patenting per capita measure for outliers likely due to errors in location coding by the USPTO, finding a few instances in which manual correction was possible. However, to guard against the possibility that any miscoding remains, we winsorize the resulting distribution of the change in patents per capita outcome variable.

cation. The final measure of interest for county d in period t is the change in the growth rate skewness measure over the five years from $t-1$ to t.

Other Data. We measure local income using average annual wage bills computed from the Quarterly Census of Wages (QCEW) dataset provided by the US Bureau of Labor Statistics. The data stem from state-level unemployment insurance records. It records employment and wages at the county-by-industry-by-year level starting in 1975. We compute the total wages per capita in a given county-year combination and use the Personal Consumption Expenditure Price Index to adjust wages. The outcomes of interest in specifications studying income growth is the change in wages per capita in county d over the five-years period ending in t.

Summary Statistics. Table 1 reports summary statistics on the outcomes described above, as well as various other instruments and derived variables studied below. The series are observed at the county by 5-year window level. The table reveals sensible patterns. Counties on average received around 1.4 thousand non-European immigrants in each 5-year period between 1975 and 2010, a meaningful contribution to overall population growth of around four thousand. Innovation (as measured by per capita patenting) increased on average over the period, with substantial heterogeneity across counties. As emphasized by the dynamism literature, measures of creative destruction including job creation rates, job destruction rates, and growth rate skewness declined on average during our sample, although the average obscures wide differences in experience: some counties became substantially more dynamic over the period we study. Wages per capita grew on average, as expected. The statistics on the remaining variables, reflecting the variation in subsets of our data or several constructed instruments, will become useful in our discussion below.

3 Constructing a Valid Instrument for Immigration

Our aim is to estimate the causal impact of immigration on innovation and growth. To do so, we estimate the following equation

$$\Delta Y_d^t = \delta_t + \delta_s + \beta \cdot \text{Immigration}_d^t + \epsilon_d^t, \tag{1}$$

where Immigration_d^t is a measure of the number of migrants flowing into destination county d at time t, ΔY_d^t is a change in either a measure of innovation or of local dynamism, and δ_t and δ_s are full sets of time and state fixed effects.

The main concern with a simple OLS estimate of (1) is that unobserved factors may affect both immigration and innovation or dynamism. For instance, it is likely that migrants are disproportionately drawn to more innovative destinations within the US.

One possibility would be to construct a "shift-share" instrument in the spirit of Altonji and Card (1991) and Card (2001), predicting immigration flows using the interaction of pre-existing ancestry shares in a destination county with current inflows from origin countries, and then summing over origin countries. However, it appears likely that omitted factors which make a set of US counties more innovative may also have attracted disproportionately many migrants from specific sets of origin countries in the past, rendering pre-existing ancestry shares endogenous. For example, Indian engineers may have historically migrated to Silicon Valley and to other information technology hubs because those provided attractive employment opportunities; and more Indian engineers may systematically migrate to Silicon Valley and other information technology hubs whenever there is a boom in software development. If this were indeed the case, the canonical shift-share approach outlined above would falsely identify a causal effect of immigration on innovation, when in reality innovations in software are the reason why destinations with high pre-existing Indian ancestry shares receive more immigration. Thus, if ancestry shares are themselves endogenous – i.e. are potentially correlated with unobserved factors affecting innovation – this poses a challenge to the canonical shift-share approach (Goldsmith-Pinkham, Sorkin, and Swift, 2018).

To overcome this challenge we augment the canonical shift-share approach with a set of instruments that isolate quasi-random variation in the pre-existing ancestry composition of US counties. This variation results only from the coincidental timing of two historical forces driving migration to the US: (i) time-series variation in the relative attractiveness of different destinations within the United States to the average migrant arriving at the time (e.g. end of nineteenth century Midwest versus early twentieth century West) and (ii) the staggered arrival of migrants from different origins (e.g. end of nineteenth century China versus early twentieth century Japan). We use the interaction of these two forces to instrument for the distribution of ancestries across US counties, and then use those exogenous ancestry shares to predict migrations. Doing so, we are immune to concerns that pre-existing ancestry shares may be themselves endogenous.

3.1 Constructing an Instrument for Immigration

In order to construct a valid instrument for immigration, we build upon the methodology developed in Burchardi et al. (forthcoming) combined with the "shift-share" approach of Altonji and Card (1991) and proceed in three steps. While this procedure may seem complex when considered in its entirety, each step is easy to implement, and follows a simple logic. In a nutshell, we first use the quasi-random timing of the interaction between the arrival of migrants from different origin countries to the US, and the relative attractiveness of destination counties to all migrants. Summing over cohorts of migrants, this allows us to predict the number of residents of a given US county with ancestry from a given origin country in a baseline year. Second, we interact those predicted pre-existing ancestry levels in the baseline year with the overall number of migrants arriving from a given origin country in the five years after that baseline year, to predict the number of migrants from a given origin to a given US county during that period. Finally, we sum this predicted number of recent migrants across all origins for each US county, to get a valid instrument for the overall inflow of migrants from all origins into a given US county. Formally, the three steps of our procedure are as follows.

Step 1: Isolating quasi-random variation in ancestry. We predict the number of residents of destination county d with ancestry from origin country o in baseline year t (in thousands), $A^t_{o,d}$, by estimating

$$A^t_{o,d} = \delta^t_{o,-r(d)} + \delta^t_{c(o),-d} + X'_{o,d}\beta^t + \sum_{\tau=1880}^{t} b^\tau_{r(d)} \tilde{I}^\tau_{o,-r(d)} \frac{I^\tau_{Euro,d}}{I^\tau_{Euro}} + u^t_{o,d}, \tag{2}$$

$$\text{with } \tilde{I}^\tau_{o,-r(d)} = I^\tau_{o,-r(d)} \frac{I^\tau_{Euro,r(d)}}{I^\tau_{Euro,-r(d)}}.$$

$I^t_{o,-r(d)}$ is the total number of migrants arriving from o who settle in locations *outside* of the region where d is located over the 5-year period ending in t. $I^t_{Euro,d}$ is the number of is the total number of migrants from Europe who settle in d at time t, and $I^t_{Euro} = \sum_d I^t_{Euro,d}$. Further, $I^t_{Euro,r(d)}$ (respectively $I^t_{Euro,-r(d)}$) is the total number of migrants from Europe who settle in the region where d is located (resp. outside this region) in the period ending in year t. $\delta^t_{o,-r(d)}$ and $\delta^t_{c(o),-d}$ are a series of origin country-destination region and continent of origin-destination interacted fixed effects, while $X_{o,d}$ contains a series of time invariant controls for $\{o, d\}$ characteristics (including distance and latitude difference). We estimate (2) separately for each $t = 1980, 1985, 1990, 1995, 2000, 2005, 2010$ on all non-European countries.

From this estimation, we derive predicted ancestry in county d from origin o at time t as

$$\hat{A}^t_{o,d} = \sum_{\tau=1880}^{t} \hat{b}^{\tau}_{r(d)} \left(\tilde{I}^{\tau}_{o,-r(d)} \frac{I^{\tau}_{Euro,d}}{I^{\tau}_{Euro}} \right)^{\perp}, \tag{3}$$

where $\left(\tilde{I}^{\tau}_{o,-r(d)} \frac{I^{\tau}_{Euro,d}}{I^{\tau}_{Euro}} \right)^{\perp}$ are residuals of a regression of $\tilde{I}^{\tau}_{o,-r(d)} \frac{I^{\tau}_{Euro,d}}{I^{\tau}_{Euro}}$ on $\delta^t_{o,-r(d)}$, $\delta^t_{c(o),-d}$ and $X_{o,d}$.

This procedure for identifying plausibly exogenous variations in ancestry builds upon Burchardi et al. (forthcoming). We implicitly assume that two forces are present in motivating the choice of where foreign migrants settle in the US: a "push" factor, the total number of migrants from country o that settle in the US in the five-year period denoted by t; and a "pull" factor, the proportion of all European migrants to the US that settle in county d in period t. The intuition is as follows. More migrants from origin country o will tend to settle in any location in the US in periods when many migrants happen to arrive from this country o in aggregate: when $I^{\tau}_{o,-r(d)}$ (the push factor) is large in some period τ. And those migrants will have a tendency to disproportionately settle in a specific destination county d if this county is precisely at the same time attractive to migrants in general, irrespective of their origin: for instance if $\frac{I^{\tau}_{Euro,d}}{I^{\tau}_{Euro}}$ (the pull factor) is large in the same period τ. Intuitively, if coincidentally a county d is attractive to migrants of all origins precisely at the time when many migrants arrive from country o, we expect many of the migrants from o to settle in d. It is this variation in the stock of d's ancestry from o which we exploit to create exogenous variation in the ancestry composition of county d at any time t. Summing over many periods, equation (2) cumulates this variation in the stock of residents with a given ancestry over repeated waves of migrations starting before 1880.

To further guard against confounding factors, we combine this "push-pull" approach with a stringent leave-out restriction. Our focus in the main regressions of interest is on migrants who arrived in the US in recent decades, since 1970, from non-European countries. This is a period over which indeed most migrants were *not* coming from Europe. We predict the historic settlement patterns of non-European immigrants, and hence their associated modern ancestry stocks, from a pull factor based on *nothing but* European migrants. This ensures that our results are not driven by origin countries with similar characteristics and therefore similar settlement patterns. In addition, for the push factor we take all migrants from o arriving to the US and remove those migrants that happen to settle in the census division where d is located. This exclusion removes the possibility that the time variation in the total number of migrants from o to the US is being driven by changes in migrant flows to the census division in which d is located. Put differently, any unobserved time-varying factors which make destinations in the census division $r(d)$ particularly attractive for migrants from o, and hence attract immigration from o to that census region, will not confound our measure of the push factor of o to $r(d)$.[7] Moreover, to adjust for the fact that some divisions may be larger than others, we add a simple scaling factor $\frac{I^{\tau}_{Euro,r(d)}}{I^{\tau}_{Euro,-r(d)}}$.

Finally, as we are predicting ancestry at the $\{o, d\}$ level, we are able to control for a very large set of fixed effects as well as $\{o, d\}$-specific controls. Any factor which would make a given destination county more attractive to migrants from a specific continent, or make migrants from a given origin settle in a specific region, is controlled for by the set of $\delta_{o,-r(d)}$ and $\delta_{c(o),-d}$ fixed effects. We control for the distance between origin country o and destination d to account for variation in migration due to proximity and we control for latitude distance between o and d to account for the potential effect of "similar climates" on migration choices. Notice that in (3) we only use variation in push-pull factors that is orthogonal to any of those potentially confounding factors.

To fix ideas around these push and pull factors more concretely, directly examining the underlying variation proves useful. Figure 1 plots the share of non-European immigration into the US from the 14 non-European origin nations with the largest cumulative immigration to the US. This push-factor variation within countries is generally clustered in time in "bursts" of immigration to the US, typically driven by historic events in the home countries. For example, Mexican migration to the US experiences a spike during the period of the Mexican Revolution from 1910-20. Cuban immigration flows increase during the 1960s and 1970s in the decades after the Cuban revolution. Immigration from Vietnam reaches substantial numbers only from the mid-1970s onwards in the wake of US involvement in the Vietnam War. Chinese and Japanese migration to the US fell from relatively higher levels early in the sample to low levels before rising over time, in this case as the various US immigration exclusion acts were repealed.

As the different nations in Figure 1 sent immigrants at different points in time to the US, the location of relatively attractive destination counties – our source of variation for the pull factor – shifted substantially. Figure 2 plots color-coded maps of migration into the US over Census waves from 1880 to 2010, with darker shades representing a higher intensity of migration to a given county. Early on in the sample during the late 19th century northeastern locations were particularly attractive destinations. But by the early 1900's the favored destinations shifted to the midwest and western regions, before shifting yet again to the coastal and southeastern regions later on.

[7]Note that our approach here imposes a stricter requirement than simply excluding the number of migrants to d.

So to summarize, the rich variation in Figures 1 and 2 allows us to consider a narrow subset of migration flows and hence resulting ancestry levels. Instead of, say, considering data on all flows of Mexican migrants to the US, our eventual set of quasi-random variation instead exploits, say, the fact that certain southeastern and midwestern regions happened to be popular destinations ("pulling" people into destination counties d) during the period of heavy Mexican immigration around the Mexican Revolution ("pushing" people out of that origin nation o).

Step 2: Predicting migrations from individual countries. Having isolated plausibly exogenous variations in the stock of ancestry at the $\{o, d\}$ level for all periods after 1970, we are now ready to implement a "shift-share" method similar to Altonji and Card (1991), without being worried that ancestry itself is an endogenous variable. We predict immigration from o to d in period t by estimating

$$I_{o,d}^t = \delta_{o-r(d)} + \delta_{c(o)-d} + X_{o,d}'\beta + \sum_\tau \gamma^\tau \cdot \delta_\tau \cdot [\hat{A}_{o,d}^{\tau-1} \times \tilde{I}_{o,-r(d)}^\tau] + \nu_{o,d}^t, \tag{4}$$

where $\hat{A}_{o,d}^{\tau-1}$ is predicted ancestry from (3) and δ_τ are time fixed effects. This equation is based upon the simple idea that newly arrived migrants from o ($\tilde{I}_{o,-r(d)}^\tau$) will tend to settle in places where a large community of individuals with the same o ancestry already reside ($\hat{A}_{o,d}^{\tau-1}$).

Again, to guard against confounding factors, when predicting migrations from o to d, we leave out o's migrants who settle not only in d, but also in the entire census division where d is located. As before, we control for a large set of origin-region pair and destination-continent fixed effects as well as distance and latitude-distance between o and d.

Step 3: Predicting aggregate migrations. We are finally able to generate our main instrument for the total number of migrants settling in county d in period t, Immigration_d^t in equation (1),

$$\hat{I}_d^t = \sum_o \hat{\gamma}^t \cdot [\hat{A}_{o,d}^{t-1} \times \tilde{I}_{o,-r(d)}^t]. \tag{5}$$

A sufficient condition for the validity of this instrument can be written as

$$\tilde{I}_{o,-r(d)}^s \frac{I_{Euro,d}^s}{I_{Euro}^s} \perp \epsilon_d^t \forall o, d, s \leq t. \tag{6}$$

It requires that confounding factors that correlate with increases in a given county's innovation or dynamism post-1975 and historically made a given destination more attractive for migration from a given non-European origin country do not also correlate with past instances of the interaction of the settlement of European migrants with the total number of migrants arriving from that non-European origin who settle in other US census regions. If this condition is satisfied, the ancestry shares used to predict immigration in Step 2 are exogenous, as is the variation in total immigration calculated in (5).[8]

We believe this assumption is plausible: consider again a shock to productivity of software development in Silicon Valley that attracts Indian software engineers. This confounding shock, and any other origin-destination specific factor that drives migration and might affect the destination's capacity for future innovation, in general affects neither $\hat{A}_{o,d}$ nor $\tilde{I}^t_{o,-r(d)}$; the former depends only on how the historical destination choices of Europeans coincided with the number of Indians arriving in the US who chose destinations other than the West Coast, while the latter depends only on the number of Indians arriving in t who again do not settle on the West Coast. In order to violate (6), the confounding shock would instead have to systematically affect both the destination choices of Indians and large numbers of Europeans (enough to sway shares), while also attracting large numbers of Indians to US counties outside of the West Coast.

3.2 The Construction and Performance of the Instrument

We now review the estimation results of each of the steps towards the construction of our instrument, including the performance of the resulting instrument for county-level immigration in the relevant first-stage regression.

In step 1 of our instrument construction, we predict ancestry levels by using historical push-pull factors in (3). Figure 3 reports the coefficients in this regression predicting 2010 ancestries, and reports the coefficients on the interaction term by time period (assuming for presentational purposes only that $b^{\tau}_{r(d)} = b^{\tau}\ \forall r(d)$). The results indicate that we identify variation in current ancestry levels based on push and pull factors from across the full range of time periods in our sample, with precise contributions from periods as far back as the pre-1900's census waves. These coefficients are positive and mostly significant. The negative coefficient in the late 1920s is consistent with large return-migrations during the Great Depression, when arriving migrants swiftly returned home and possibly attracted earlier migrants to follow their steps. Figure 4 presents a bin scatter plot of the resulting predicted ancestry levels against realized ancestry in 2010. A strong link is evident between realized ancestry and predicted ancestry along the

[8]Exogeneity of ancestry shares is generally not a necessary condition for the validity of the canonical shift-share approach, and is not a necessary condition for our approach either. For work identifying necessary and sufficient conditions for the validity of the shift-share instrument as proposed by Altonji and Card (1991) and Bartik (1991) see Borusyak, Hull, and Jaravel (2018).

45-degree line. We conclude that Step 1 successfully predicts plausibly exogenous variation in ancestry.

In step 2 of our instrument construction, we interact lagged predicted ancestry with contemporaneous scaled push factors ($\hat{A}_{o,d}^{t-1} \times \tilde{I}_{o,-r(d)}^{t}$) to predict plausibly exogenous variation in immigration $I_{o,d}^{t}$ at the $\{o, d\}$ level in (4). We allow the coefficient γ to vary by time period t, and Table 2 reports the resulting estimates. Our ancestry and push factors positively predict immigration at the origin-county level in all time periods, with R^2 values indicating high explanatory power in Column 1 with no other predictors included. Adding a range of other controls and fixed effects in Columns 2-4 changes the overall R^2 values by little, and the coefficients on ancestry and push factors remain stable even in Column 5 when including the endogenous total flow of European migration to the same county as an additional control. We conclude that Step 2 successfully predicts plausibly exogenous variation in immigration at the $\{o, d\}$ level.

In step 3, we sum across origin nations to compute an instrument $\hat{I}_{d}^{t}$ for total non-European immigration to county d at time t (5). Table 3 reports the first-stage regression of county-level non-European immigration on this instrument. Column 1 is the simplest specification, regressing immigration on the instrument itself alone, with a strong first-stage F-statistic and a positive impact of our instrument on realized immigration. Column 2 allows for census division level and time fixed effects, with a stable coefficient and continued first-stage strength. The same conclusion is reached with disaggregate state and time effects in Column 3. Later on we also analyze a more restricted sample of counties for use with the natively MSA-level BDS data described above, and Column 4 shows that a strong positive first-stage relationship exists in this narrow subsample as well.

4 The Impact of Immigration on Innovation and Growth

Classical endogenous growth theories links population growth to innovation, dynamism, and income growth at the local level, a result which we lay out in a straightforward theoretical derivation in Appendix B. In this section, we exploit our quasi-random variation in immigration above to test these predictions explicitly.

4.1 Immigration and Innovation

We first test the hypothesis that immigration *causes* an increase in innovation at the county level. Table 4 shows estimates of (1) where we instrument for the number of immigrants arriving in the county during the 5-year period. The dependent variable is the change in patents per capita over the same period. Column 3 shows our standard specification which includes state and time fixed effects, thus controlling for differential trends in innovation growth at the state level. The estimated effect is positive and statistically highly significant (0.035, s.e.=0.012). It implies that the arrival of 1000 additional immigrants in a given county increases the number of patents filed annually per 1,000 inhabitants by 0.035 relative to the previous 5-year period. In other words, $1/0.035 \times 1,000 \approx 30$ thousand more immigrants cause an increase of 1 more patent per 1,000 on average. Comparing these magnitudes to the summary statistics above, we see that an increase in immigration flows of one standard deviation - 12 thousand immigrants - causes around 0.4 more patents per 1,000 people, an increase of 21% relative to mean.

The remaining columns show variations of the IV estimate that use different sets of fixed effects. Our preferred estimate in Column 3, controlling for highly disaggregated state time trends and common time effects, is conservative.

Robustness: Alternative Instruments. Table 5 shows how our instrumentation and identifying assumptions affect these results. Column 1 shows the OLS estimate of (1) for comparison. As expected, it is somewhat larger than our preferred estimate (by about 1.5 standard errors), consistent with the view that, other things equal, immigrants select into innovative counties in equilibrium. In Column 2 we use a traditional shift-share approach where we skip steps 1 and 2 of our procedure and instead directly instrument immigration to county d in period t with $\sum_o A_{o,d}^{t-1} \times I_o^t$, where here $A_{o,d}^{t-1}$ represents baseline ancestry in shares as opposed to levels. Column 3 shows a slightly more sophisticated approach where we skip only step 1 and use $A_{o,d}^{t-1} \times I_o^t$ directly in (4), again using baseline ancestry shares. The former approach would yield consistent estimates if there were no unobserved characteristics of destinations that drive both immigrations from all origins and a county's capacity to innovate. This assumption is almost certainly violated. For example, it is likely that a local university might attract students from many parts of the world, and that some of these students may stay in the area after graduation and innovate. The latter approach is somewhat more stringent. It would yield consistent estimates if the interaction $A_{o,d}^{t-1} \times I_o^t$ were exogenous conditional on the rich set of fixed effects in (4); however, it nevertheless requires that there be no unobserved origin-destination specific factors, akin to the productivity shock in our example of Indians in Silicon Valley above, that affect both past migrations and a county's capacity to innovate. Column 4 shows our standard specification for comparison, which controls for such confounding factor.

Comparing the coefficients across Columns 1-4 we find that all four are positive and statistically highly significant. While the differences across estimates are not very large, ranging from 0.055 to 0.035, each tightening of the identifying assumptions is nevertheless associated with a drop in the coefficient of interest. This suggests that the forms of endogenous selection mentioned above do operate in the data, although they are not strong enough to overturn the overwhelmingly positive association between immigration and measured innovation in our baseline estimate in Column 4. We also examine two alternative approaches to our baseline algorithm for constructing the instrument outlined in Steps 1-3 above. The first is to substitute predicted ancestry counts $\hat{A}^t_{o,d}$ with the prediced ancestry shares (constructed from those predicted counts), with the result in Column 5. The second is to use a different leave-out approach in Step 1, only exploiting "push factors" for a given destination county d that reflect flows of immigrants with the same origin o to other counties whose overall immigration flows are uncorrelated with flows to d over the full sample. The resulting estimates are reported in Column 6. In both cases, we see that a positive impact of immigration on innovation survives, and our baseline results in Column 3 remain conservative.

Robustness: Alternative Samples. Table 6 probes the robustness of these results by dropping important origin countries and destination counties. The columns in Panel A show estimates of our standard specification while dropping migrations from the five largest sending countries post 1975 (Mexico, China, India, Philippines, and Vietnam) from the sum in (5), thus treating migrations from these countries as endogenous. While dropping Mexican migrations from our instrument lowers the F-statistic in the first stage by about half, the coefficient estimates are almost unchanged across these variations, thus showing that no single large sending country drives our results.

Panel B similarly drops possibly influential destination counties from the panel. Columns 1 through 3 drop the top 10 counties in terms of total immigration between 1975 and 2010; population in 2010; and growth in patents per capita between 1975 and 2010, respectively. Columns 4 and 5 then drop Los Angeles County and New York County from the panel, respectively. In the case of dropping large population centers or immigrant destinations (including Los Angeles County) the coefficient increases significantly; this reflects the fact that innovation varies with changes in immigration and does not necessarily occur disproportionately in the most populated counties. We continue to see meaningful and positive impacts of immigration on innovation across those restricted samples, with similar coefficients when dropping New York county or the 10 most innovative counties.

Robustness: Additional Controls. In Table 7 we go one step further and control parametrically for a number of initial conditions that could be considered drivers of long-term economic growth: population density in 1970, the number of patents generated in 1975 per 1,000, 1970 inhabitants (1975 is the first year for which our patent data is available), and the share of the 1970 population that is high-school and college educated, respectively. All of these covariates could be considered "bad controls" (Angrist and Pischke, 2009), in the sense that they are themselves outcomes of migration and should thus more appropriately be thought of as various channels through which historical migrations and ancestry may affect innovation. Nevertheless, it is comforting for our identifying assumption that controlling for these initial conditions has only modest effects on our result. The largest change in the coefficient of interest occurs when we include the share of the population with a college education in 1970, lowering it by slightly less than one standard error to 0.026 (s.e.=0.010) (or 0.024 (s.e.=0.009) when including all controls). Throughout these variations, the estimated effect of immigration remains positive and statistically highly significant.

4.2 Immigration and Economic Dynamism

In Table 8, we supplement our analysis of immigration's impact on innovation with a range of additional economic dynamism and income growth outcomes which endogenous growth theory suggests should link positively to innovation. The job creation and destruction rates computed from the BDS data narrow the sample to its smallest level. So we replicate our main specification of the change in patents per capita on immigration for this restricted sample in Column 1, where we still observe a positive and similarly sized impact of immigration on innovation.

Immigration causes an increase in creative destruction or gross flows in jobs, as reported in Columns 2 and 3. Both the job creation rate and job destruction rate increase in the face of more immigration, with the implication that overall churning or reallocation in the labor market responds positively. Recall that dynamism measures decline on average over this period, as emphasized by the wide literature on declining creative destruction in the US. Our positive estimated responses to immigration indicate that immigration may help to dampen such declines. Turning to magnitudes, a one standard deviation increase in immigration in a county - around 12 thousand more people - leads to an increase of job creation rate of 2.1 percentage points (around 7% relative to the mean decline) and an increase in the job destruction rate of 1.8 percentage points (around 11% relative to the mean decline).[9]

Exploring our alternative measure of dynamism, higher immigration causes an increase in the skewness of employment growth (see Column 4). Intuitively, when more immigrants arrive, the importance of superstar employment growth experiences across sectors in an area increases. A one standard deviation increase in immigration causes about a 3% increase in skewness relative to the mean decline in this measure over the sample.

4.3 Immigration and Income Growth

At their core, endogenous growth models link innovation to income growth, and Column 5 of Table 8 reports that more immigration causes an increase in wages per person. Immigration of around 12 thousand more people to a county increases wages per capita by around 5% relative to the mean observed growth.

To summarize the estimates in this section, immigration causes moderately large increases creative destruction and income growth at the local level, validating traditional endogenous growth theories and potentially serving as a potent counterweight to trend decline in dynamism and growth in the US in recent decades.

5 Spillovers and Education

The local positive impact of immigration on innovation that we document above validates long-standing theoretical mechanisms linking innovation to growth in population. However, two natural questions remain. First, if ideas and goods flow across regions, to what extent do the impacts of immigration spill over across counties in our data? Second, since most theoretical models predict that more highly skilled immigrants bring more effective input to bear for innovation or production, to what extent do the impacts of immigration on innovation vary with the education level of migrants? We tackle both issues directly in this section, finding that positive spillovers appear meaningful and that the impact of immigration on innovation increases with average schooling levels.

[9] It is worth noting that although most endogenous growth theories link higher dynamism to innovation, higher income growth, and higher welfare, the impact of dynamism on the subjective wellbeing of individuals exposed to such creative destruction is more ambiguous (Aghion et al., 2016).

5.1 Spatial Spillovers

To explore the impact of cross-county spillovers, we consider two geographic spillover concepts in Table 9. First, we consider within-state spillovers, constructing for each destination county d at each time t a measure of immigration to all counties other than d in the same state. This measure, labeled $I^t_{S,nonEuro}$, varies at the same level as the county-specific baseline immigration flow $I^t_{d,nonEuro}$. We face a challenge of developing an instrument for this second endogenous variable. But recall that our original instrument $\hat{I}^t_d$ provides a county-level measure of immigration shocks. So to construct an instrument for the state-level immigration flows we follow a symmetric procedure, adding the immigration shocks up for all other counties within the same state as d.

In a second approach to spillovers relying on a different notion of geography than state boundaries, we consider a specification allowing spillovers from county-to-county to vary smoothly by distance. For any county d at each time t, we construct the sum of all immigration to other counties in the same census division, inversely weighted by the distance to the reference county d. The distance measures reflect a matrix of great circle distances computed from county centroids using the Census mapping files for county geographies. The resulting distance-weighted measure of immigration to other counties, labeled $I^t_{IDW,nonEuro}$, varies at the county d by time t level. As in the case of state-level spillovers, we symmetrically sum in an inverse distance weighted fashion across other counties' predicted immigration shocks to construct an instrument for the distance-based spillover immigration variable.

In the first column of Table 9, we report a duplicate IV estimate of the effect of own-county immigration on innovation. The second column adds a second endogenous variable, the state-level sum across other counties' immigration. The first-stage F statistics reveal strong power for both the own-county and state-level immigration flows.[10] The impact of own-county migration on immigration remains strongly positive with a similar magnitude, but more immigration to other counties within the same state also increases innovation. The magnitudes implied by Column 2 are sensible. A one-standard deviation increase in immigration to your own county (12 thousand people) increases patenting per capita by 20.8% relative to mean, holding state-level immigration to other counties constant. Similarly, a one-standard deviation increase in immigration to all other counties in the state (1.4 million more immigrants), holding your own county's immigration flow constant, increases patenting per capita by around 19.8% relative

[10] For all specifications involving multiple endogenous variables, we use the Angrist and Pischke first-stage F-statistic, testing for each regressor separately the null of weak identification.

to mean. In other words, both immigration locally and immigration to your state positively impact your own innovation. Although migrants to other counties matter less individually for an individual county's own innovation, the larger scale of those flows means that such immigrants bring similarly sized innovation impacts to the local economy.

Column 3 moves on from the state-level spillovers analysis to analyze the impact of the inverse-distance weighted measure of immigration to other counties on local innovation. The first-stage F-statistics again reveal strong power separately for local immigration and the spillovers measure. The impact of your own local immigration flow remains positive and of a similar magnitude, while immigration to other counties nearby according to the inverse-distance weighted criterion also cause an increase in local innovation. The magnitudes of the coefficients bear further explanation again. The local impacts are similar to the results from Column 2. A one-standard deviation increase in immigration to your own county (12 thousand people) increases patenting per capita by 17.3% relative to mean, holding state-level immigration to other counties constant. But the arguably more coherent measure of proximity based on distance rather than state boundaries reveals stronger spillover impacts. A one standard deviation increase in immigration to all other nearby counties according to the inverse-distance criterion increases local patents per capita by 50% relative to mean. So while we continue to estimate strong positive impacts of local immigration flows on local innovation, individual counties' innovative activity benefits in a large fashion from immigration to their region, broadly defined, as well.

5.2 Education of Immigrants

We now explore whether more educated immigrants have different impacts on innovation. First, to measure educational attainment for individuals who might reasonably have had the time to attend school, we limit ourselves to the analysis of immigrants age 25 or greater, constructing the endogenous measure of immigration at the county level within this subset of immigrants. We then interact this overall adult immigration flow with the average schooling levels - total years of education or total years of college, in two alternative versions - adding a second endogenous variable to our baseline specification.

To successfully instrument for both the total immigration flow and the interaction of immigration and education, we exploit the fact that in our initial instrument construction we created quasi-exogenous immigration shocks for each origin country-o $\times$ destination county-d pair in each time period t, $\hat{I}^t_{o,d} = \hat{A}^{t-1}_{o,d} \times \tilde{I}^t_{o,-r(d)}$. We disaggregate our baseline instrument to this level, using the predicted immigration shocks for each of the the top 20 origin nations as a joint set of instruments for both total immigration and immigration interacted with average education. Because different nations have different schooling levels, their immigrants naturally also have different average schooling attainment. Intuitively, our estimation strategy exploits variation across counties with similar total immigration levels but which receive flows from different nations, some of which are predictive of immigration with a high average schooling and some of which are predictive for immigration with a lower average education.

Table 10 reports the results of our analysis. Column 1 replicates our core specification for the age 25+ immigration sample, continuing to reveal a positive - now slightly stronger than baseline - impact of immigration on the growth of patenting per capita. Column 2 adds the interaction of immigration with (demeaned) average years of education for immigrants to the same county. First-stage F-statistics reveal the strong power of our instrumentation strategy for both total immigration and the interacted immigration measure. The estimates indicate that more highly educated immigrants cause a bigger increase in innovation. To inspect the magnitude of the heterogeneity at work here, we first split the distribution of schooling levels into deciles. Figure 5 plots the impact of immigration on patenting per capita at each of these deciles of schooling implied by the estimates in Column 2. Moving from the 30-th %ile level of schooling of immigrants across counties to the 80-th %ile increases the impact of 1000 extra immigrants on innovation from around 0 to around 0.4 patents per 1000 people. In other words, while a county receiving lower-educated immigrants would see an essentially neutral impact on innovation, a county receiving highly skilled immigrants receives an additional patent per 1,000 for every extra $1/0.4 \times 1,000$ = 2,500 immigrants. In other words, the effect of skilled migrants (at the 80-th percentile of the education distribution of migrants) is almost one order of magnitude stronger than the effect of the average migrants. Column 3 reports a similar analysis, but measuring educational attainment by average years of college completed rather than average years of total schooling, and Figure 6 again plots the marginal impact of immigration on innovation by deciles. Unsurprisingly, the impact of immigration on innovation increases even more strongly with college attainment than with overall educational attainment.

6 Conclusion

We find that plausibly exogenous variation in migration at the county level induced by quasi-random variation in historical "push" and "pull" factors for immigration strongly predicts overall realized immigration. We then estimate that over the past four decades immigration at the local level has spurred increased innovation or patenting per capita. A strong tradition in endogenous growth theory predicts exactly these links, with more immigrants bringing new ideas, purchasing power, and scale which contribute to innovation. The impact of immigration on innovation spills over positively to nearby counties and unsurprisingly increases with the schooling of immigrants.

However, we also find that immigration increases dynamism or creative destruction at the local level. Classic theoretical models predict that such creative destruction, which has been declining on average in the US overall in recent decades, should move with overall innovation, consistent with our findings on patents above. We conclude that increased immigration in recent decades may have prevented even worse declines in dynamism at the local level, and might contribute to further mitigating the downward trend in dynamism if allowed in the future.

Finally, we show that immigration to an area also has increased income growth or wages per capita at the local level.

Given the relevance of innovation, dynamism, and growth for the long-run prospects of the US economy as well as for macro models of growth, we view our results as relevant to the ongoing debate about the appropriate choice of immigration policy.

References

Aghion, P., U. Akcigit, A. Deaton, and A. Roulet (2016): "Creative destruction and subjective well-being," *American Economic Review*, 106, 3869–97.

Aghion, P. and P. Howitt (1992): "A model of growth through creative destruction," *Econometrica (1986-1998)*, 60, 323.

Akcigit, U., J. Grigsby, and T. Nicholas (2017): "The rise of American ingenuity: Innovation and inventors of the Golden Age," National Bureau of Economic Research working paper.

Alon, T., D. Berger, R. Dent, and B. Pugsley (2018): "Older and slower: The startup deficit's lasting effects on aggregate productivity growth," *Journal of Monetary Economics*, 93, 68–85.

Altonji, J. G. and D. Card (1991): "The effects of immigration on the labor market outcomes of less-skilled natives," in *Immigration, trade, and the labor market*, University of Chicago Press, 201–234.

Angrist, J. D. and J.-S. Pischke (2009): *Mostly harmless econometrics: An empiricist's companion*, Princeton university press.

Bartik, T. (1991): *Who Benefits from State and Local Economic Development Policies?*, W.E. Upjohn Institute.

Bernstein, S., R. Diamond, T. McQuade, and B. Pousada (2018): "The Contribution of High-Skilled Immigrants to Innovation in the United States," Working paper.

Bloom, N., C. I. Jones, J. Van Reenen, and M. Webb (2017): "Are ideas getting harder to find?" National Bureau of Economic Research working paper.

Borusyak, K., P. Hull, and X. Jaravel (2018): "Quasi-Experimental Shift-Share Research Designs," Working Paper.

Burchardi, K., T. Chaney, and T. Hassan (forthcoming): "Migrants, Ancestors, and Foreign Investments," *Review of Economic Studies*.

Card, D. (2001): "Immigrant Inflows, Native outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, 19, 22–64.

Decker, R., J. Haltiwanger, R. Jarmin, and J. Miranda (2014): "The secular decline in business dynamism in the US," Working paper.

Duncan, B. and S. J. Trejo (2017): "The complexity of immigrant generations: Implications for assessing the socioeconomic integration of Hispanics and Asians," *ILR Review*, 70, 1146–1175.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2018): "Bartik Instruments: What, When, Why and How," Working Paper.

Gordon, R. J. (2012): "Is US economic growth over? Faltering innovation confronts the six headwinds," National Bureau of Economic Research working paper.

Grossman, G. M. and E. Helpman (1991): "Quality ladders in the theory of growth," *The Review of Economic Studies*, 58, 43–61.

Hathaway, I. and R. E. Litan (2014): "Declining business dynamism in the United States: A look at states and metros," *Brookings Institution*, 2.

Hopenhayn, H., J. Neira, and R. Singhania (2018): "From Population Growth to Firm Demographics: Implications for Concentration, Entrepreneurship and the Labor Share," National Bureau of Economic Research working paper.

Jones, C. I. (1995): "R&D-Based Models of Economic Growth," *Journal of Political Economy*, 103.

——— (1999): "Growth: with or without scale effects?" *American Economic Review*, 89, 139–144.

Karahan, F., B. Pugsley, and A. Sahin (2016): "Demographic origins of the startup deficit," New York Fed working paper.

Kerr, S. P. and W. R. Kerr (2016): "Immigrant entrepreneurship," National Bureau of Economic Research working paper.

Klette, T. J. and S. Kortum (2004): "Innovating firms and aggregate innovation," *Journal of Political Economy*, 112, 986–1018.

Peretto, P. F. (1998): "Technological change and population growth," *Journal of Economic Growth*, 3, 283–311.

Peters, M. (2017): "Refugees and Local Agglomeration – Evidence from Germany's Post-War Population Expulsions," Working paper.

Romer, P. M. (1990): "Endogenous technological change," *Journal of Political Economy*, 98, S71–S102.

Sequeira, S., N. Nunn, and N. Quian (forthcoming): "Immigrants and the making of America," *Review of Economic Studies*.

Solow, R. M. (1956): "A contribution to the theory of economic growth," *The Quarterly Journal of Economics*, 70, 65–94.

Tabellini, M. (2018): "Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration," Working paper.

Young, A. (1998): "Growth without scale effects," *Journal of political economy*, 106, 41–63.

Tables and figures

Table 1: Summary Statistics for County-Year Observations

	mean	sd	iqr
Panel A: Population and Immigration			
Non-Euro. immigration (1000s)	1.42	12.21	0.22
Instrument for Non-Euro. immigration (1000s)	0.00	4.99	0.24
Difference in population (1000s)	4.03	19.56	2.56
Non-Euro. immigration, age 25+	0.80	6.91	0.11
Non-Euro. immigration, age 25+: Average years college	1.56	1.40	1.82
Non-Euro. immigration, age 25+: : Average years education	11.28	3.05	4.38
State (minus own county) Non-Euro. immigration (1000s)	808.00	1,438.90	557.47
Instrument for state (minus own county) Non-Euro. immigration (1000s)	31.92	366.25	60.89
Inverse-distance weighted Non-Euro. immigration (1000s)	1.15	0.78	0.65
Instrument for inverse-distance weighted Non-Euro. immigration (1000s)	-0.01	0.15	0.03
Panel B: Innovation			
Difference in patents (per 1,000 people)	2.18	6.53	1.28
Panel C: Dynamism			
Difference in job creation rate	-32.47	209.90	50.00
Difference in job destruction rate	-17.37	199.58	38.46
Difference in skewness of employment growth	-6.82	48.91	51.87
Panel D: Wages			
Difference in average annual wage (per 1,000 people)	1.90	5.62	2.58

Notes: This table reports the mean, standard deviation, and interquatile range for the primary variables of interest. Panel A provides information on variables constructed from Census micro data on population, immigration, and education of migrants as well as instruments for immigration. Panel B provides summary statistics for our primary outcome of interest, the change in patents per 1,000 people constructed from USPTO data. Panel C then reports on the dynamism outcomes, which rely on data from the US BDS and US CBP. Finally, Panel D offers summary statistics on the wage data taken from the QCEW dataset and normalized using the Personal Consumption Expenditure Price Index.

Table 2: Regressions of Immigration on Shift-Share Instruments at the Country-County Level

	$Immigration^t_{o,d}$				
	(1)	(2)	(3)	(4)	(5)
$\hat{A}^{1975}_{o,d} \times \tilde{I}^{1980}_{o,-r(d)}$	0.0036***	0.0036***	0.0035***	0.0035***	0.0035***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\hat{A}^{1980}_{o,d} \times \tilde{I}^{1985}_{o,-r(d)}$	0.0016***	0.0016***	0.0016***	0.0016***	0.0016***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\hat{A}^{1985}_{o,d} \times \tilde{I}^{1990}_{o,-r(d)}$	0.0018***	0.0018***	0.0018***	0.0018***	0.0018***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\hat{A}^{1990}_{o,d} \times \tilde{I}^{1995}_{o,-r(d)}$	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\hat{A}^{1995}_{o,d} \times \tilde{I}^{2000}_{o,-r(d)}$	0.0004***	0.0004***	0.0004***	0.0004***	0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\hat{A}^{2000}_{o,d} \times \tilde{I}^{2005}_{o,-r(d)}$	0.0002***	0.0002***	0.0002***	0.0002***	0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\hat{A}^{2005}_{o,d} \times \tilde{I}^{2010}_{o,-r(d)}$	0.0002***	0.0002***	0.0002***	0.0002***	0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$I^t_{Euro,d}$					0.0109***
					(0.0031)
N	3,583,881	3,583,881	3,583,881	3,583,881	3,583,881
F-Stat	1.35e+06	1.36e+06	3.55e+05	3.55e+05	3.39e+05
R^2	0.656	0.657	0.709	0.709	0.709
Controls:					
Distance	no	yes	yes	yes	yes
Latitude Dis.	no	yes	yes	yes	yes
Region-Country FE	no	no	yes	yes	yes
County-Continent FE	no	no	yes	yes	yes
Time FE	no	no	no	yes	yes
Concurrent European Immigration	no	no	no	no	yes

Notes: This table reports coefficient estimates for step 2 of our instrument construction, shown in Equation (4), at the country-county level. Moving from Column 1 to Column 5 we introduce additional fixed effects and controls into the regression specification. We report the first-stage F-statistic on the excluded instruments for each specification. Standard errors are clustered by country for all specifications and *,**, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 3: Panel Regressions of Non-European Immigration on Predicted Immigration Flows at the County Level for 1980 to 2010

	5-Year Non-European Immigration ($Immigration_d^t$)			
	(1)	(2)	(3)	(4)
$\widehat{Immigration}_d^t$	2.107***	2.107***	2.100***	2.111***
	(0.046)	(0.062)	(0.061)	(0.068)
N	21,987	21,987	21,987	6,600
F-Stat	2,138.942	1,167.679	1,202.118	950.962
R^2	0.741	0.768	0.777	0.771
Controls:				
Geography FE	no	DV	ST	ST
Time FE	no	yes	yes	yes
MSA Counties	no	no	no	yes

Notes: This table reports the results for step 3 of instrument construction, or the coefficient estimates for the first stage specification. Column 1 provides the results from a regression of non-European immigration on the instrument described in equation (5). Column 2 then adds to that regression time and census division effects while Column 3, our main specification, includes state and time fixed effects. Column 4 shows the first stage estimated on a restricted sample of counties, which is used in analyses of natively MSA-level BDS data. We report the first-stage F-statistic on the excluded instrument for each specification. Standard errors are clustered by state for all specifications and *,**, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Panel Regressions of 5-year Difference in Patents per 1,000 People on Non-European Immigration at the County Level

	5-year Difference in Patents per 1,000 People for 1980 to 2010		
	(1)	(2)	(3)
$Immigration_d^t$	0.055**	0.043***	0.035***
	(0.025)	(0.015)	(0.012)
N	21,987	21,987	21,987
Controls:			
Specification	OLS	IV	IV
MSA Counties	no	no	no
Geography FE	ST	DV	ST
Time FE	yes	yes	yes

Notes: This table reports the results of our second stage specification, described in Equation (1), where the dependent variable is the change in patents per 1,000 people (population is based on baseline 1970 levels) in county d in the 5-year period ending in t and the endogenous variable is non-European immigration to d in t. Column 1 provides the results of the OLS estimation of Equation (1). Column 2 provides an IV estimate of the second stage when including time and census division controls while Column 3, our main specification, reports the results of the IV when including state and time controls. Standard errors are clustered by state for all specifications and *,**, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5: Robustness - Alternative Instruments

	5-year Difference in Patents per 1,000 People for 1980 to 2010					
Specification:	*OLS*	*Original Card*	*Advanced Card*	*Baseline Instrument*	*Share Instrument*	*Correlated Counties*
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Second Stage						
$Immigration_d^t$	0.055**	0.048**	0.041**	0.035***	0.045***	0.043***
	(0.025)	(0.019)	(0.017)	(0.012)	(0.014)	(0.010)
N	21,987	21,987	21,987	21,987	21,987	21,987
Panel B: First Stage						
$\widehat{Immigration}_d^t$		0.132***	0.985***	2.100***	17.875***	2.195***
		(0.005)	(0.027)	(0.061)	(1.035)	(0.201)
N		21,987	21,987	21,987	21,987	21,987
F-Stat		819.857	1,314.636	1,202.118	298.267	118.733
R^2		0.872	0.935	0.777	0.743	0.468
Controls:						
Geography FE	ST	ST	ST	ST	ST	ST
Time FE	yes	yes	yes	yes	yes	yes

Notes: This table displays the first stage (Panel B) and second stage (Panel A), described in Equation (1), where the dependent variable is the change in patents per 1,000 people (population is based on baseline 1970 levels) and the endogenous variable is non-European immigration to d in t. In this table, each column utilizes a different instrumentation strategy. Column 1 provides the OLS estimate of Equation (1) as a reference. Column 2 relies on a traditional shift-share procedure where the instrument is the sum over countries o of the interaction of baseline ancestry shares for country o with the number of migrants from o to the US in period t. Column 3 provides a more sophisticated approach, using the same interaction of baseline ancestry shares and contemporaneous migration but substituting this into Equation (4) to construct an instrument for non-European migration. Column 4 provides our standard specification for comparison. Column 5 then follows the same procedure as our standard instrument construction but replacing baseline ancestry in levels with predicted ancestry in shares (constructed from those predicted counts). Column 6 also follows the same approach for instrument construction as our standard instrument except that the leave-out strategy in Step 1 is changed: the push factor in this case excludes all destination counties whose overall time-series of immigration flows are correlated with those of d (as opposed to excluding counties in the same census division ($r(d)$) as d). We report the first-stage F-statistic on the excluded instrument for each specification in Panel B. Standard errors are clustered by state for all specifications and *,**, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 6: Robustness - Alternative Samples

	5-year Difference in Patents per 1,000 People for 1980 to 2010				
	(1)	(2)	(3)	(4)	(5)
Panel A Excluding:	*Mexico*	*China*	*India*	*Philippines*	*Vietnam*
$Immigration_d^t$	0.032***	0.035***	0.035***	0.035***	0.035***
	(0.010)	(0.012)	(0.012)	(0.012)	(0.012)
N	21,987	21,987	21,987	21,987	21,987
First Stage F-Stat	666.315	1,575.712	1,266.991	1,261.080	1,178.704
Panel B Excluding:	*Immigration*	*Population*	*Innovation*	*Los Angeles*	*New York*
$Immigration_d^t$	0.319***	0.346***	0.035***	0.157***	0.034***
	(0.038)	(0.027)	(0.012)	(0.027)	(0.012)
N	21,917	21,917	21,917	21,980	21,980
First Stage F-Stat	84.516	272.076	1,205.241	81.614	1,308.499
Controls:					
Geography FE	ST	ST	ST	ST	ST
Time FE	yes	yes	yes	yes	yes

Notes: This table reports the results of our second stage specification, described in Equation (1), run on alternative samples where the dependent variable is the change in patents per 1,000 people (population is based on baseline 1970 levels) and the endogenous variable is non-European immigration to d in t. Panel A drops migrations from the five largest sending countries post 1975 (Mexico, China, India, Philippines, and Vietnam) from the sum in Equation (5). In Panel B, each column utilizes the baseline instrument but removes a set of counties (or county) from the observations. Columns 1 through 3 drop the 10 counties with the highest: immigration between 1975 and 2010, population in 2010, and change in patents (the dependent variable), respectively. Column 4 then drops Los Angeles county while Column 5 drops New York county. We report the first-stage F-statistic on the excluded instrument for each specification. Standard errors are clustered by state for all specifications and *,**, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 7: Robustness - Additional Controls from Baseline Year (1970)

	5-year Difference in Patents per 1,000 People for 1980 to 2010					
	(1)	(2)	(3)	(4)	(5)	(6)
$Immigration_d^t$	0.035***	0.033***	0.034***	0.030***	0.026***	0.024***
	(0.012)	(0.011)	(0.012)	(0.011)	(0.010)	(0.009)
Population Density (1970)		0.002***				0.001***
		(0.000)				(0.000)
Patents per 1,000 People (1975)			0.199**			0.198***
			(0.078)			(0.076)
Share High School Education (1970)				15.074***		0.498
				(2.513)		(2.330)
Share 4+ Years College (1970)					46.621***	44.534***
					(7.835)	(10.559)
N	21,987	21,987	21,987	21,987	21,987	21,987
First Stage F-Stat	1,202.118	2,545.919	1,202.800	1,242.722	1,349.464	2,868.105
Controls:						
Geography FE	ST	ST	ST	ST	ST	ST
Time FE	yes	yes	yes	yes	yes	yes

Notes: This table reports the results of our second stage specification, described in Equation (1), where the dependent variable is the change in patents per 1,000 people (population is based on baseline 1970 levels) and the endogenous variable is non-European immigration to d in t. Column 1 repeats our main specification while Columns 2 through 5 add as a control county d's population density in 1970; patents per 1,000 people in 1975 (1970 population is used to match the dependent variable); share of high school educated; and share of the population with 4+ years of college, respectively. Column 6 then adds all of these controls in a single specification. We report the first-stage F-statistic on the excluded instrument for each specification. Standard errors are clustered by state for all specifications and *,**, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 8: Panel Regressions of All Outcomes on Non-European Immigration

5-Year Difference in:	*Patents per Person*	*Job Creation Rate*	*Job Destruction Rate*	*Job Growth Rate Skewness*	*Average Annual Wage*
	(1)	(2)	(3)	(4)	(5)
$Immigration_d^t$	0.025**	0.176***	0.152***	0.019***	0.008***
	(0.010)	(0.033)	(0.035)	(0.004)	(0.002)
N	6,600	6,600	6,600	12,564	21,976
Controls:					
Geography FE	ST	ST	ST	ST	ST
Time FE	yes	yes	yes	yes	yes

Notes: This table reports the results of our second stage specification, described in Equation (1), for each of our dependent variables with non-European immigration to d in t as the endogenous variable. Column 1 repeats our main specification for the change in patents per 1,000 people (population is based on baseline 1970 levels) but on the restricted sample of counties for which data is available in the MSA-level BDS survey. Columns 2 and 3 report the results of our second stage with the job creation rate and job destruction rate as the dependent variable, respectively. Column 4 then provides results for job growth rate skewness as the dependent variable while the dependent variable for the specification shown in Column 5 is the change in the average annual real wage (per 1,000 people) over the 5-year period ending in t. Standard errors are clustered by state for all specifications and *,**, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 9: Spillovers Analysis

	5-year Difference in Patents per 1,000 People for 1980 to 2010		
	(1)	(2)	(3)
$Immigration_d^t$	0.0429***	0.0371***	0.0308**
	(0.0144)	(0.0135)	(0.0125)
$Immigration_{S(d)}^t$		0.0003***	
		(0.0001)	
$Immigration_{IDW}^t$			1.4367***
			(0.2603)
N	21,987	21,987	21,987
F-Stat d	1,167.679	1,746.728	1,803.857
F-Stat Spillover		62.245	30.462
Controls:			
Geography FE	DIV	DIV	DIV
Time FE	yes	yes	yes

Notes: This table reports the results of our second stage specification with the change in patents per 1,000 people (population is based on baseline 1970 levels) as the dependent variable and non-European immigration to d in t as the endogenous variable. The first column repeats our 'baseline' specification but with census division controls as opposed to state controls to match the following specifications. Column 2 adds as a second endogenous variable total immigration to the state in which d is located ($S(d)$), excluding own-immigration to d, in period t and adds as a second instrument the comparable sum of instruments for all counties in the $S(d)$, excluding the instrument for d. Column 3 instead adds as a second endogenous variable the inverse-distance-weighted sum of immigration to all counties in the US excluding own-immigration to county d; again, we use a comparable summation to construct the second instrument for this regression, as the inverse-distance-weighted sum of our instrument for all counties in the US excluding the instrument for immigration to county d. For each specification we report the first-stage F-statistic(s), utilizing the F-statistic described in Angrist and Pischke (2009, p. 217-218) in the case of multiple endogenous variables. Standard errors are clustered by state for all specifications and *,**, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 10: Education Analysis

	5-year Difference in Patents per 1,000 People for 1980 to 2010		
	(1)	(2)	(3)
$Immigration_d^t$	0.081***	0.146***	0.200***
	(0.029)	(0.033)	(0.042)
$Immig._d^t * AvgYrEduc_d^t$		0.093***	
		(0.025)	
$Immig._d^t * AvgYrCol_d^t$			0.338***
			(0.074)
N	21,987	21,212	21,212
F-Stat (I)	2.80e+05	35786.991	7,896.593
F-Stat (Ed,Col)		9,160.908	5,614.361
R^2	0.085	0.107	0.125
Controls:			
Specification	IV	IV	IV
Geography FE	ST	ST	ST
Time FE	yes	yes	yes

Notes: This table reports the results of our second stage specification with the change in patents per 1,000 people (population is based on baseline 1970 levels) as the dependent variable and non-European immigration to d in t as the endogenous variable. Column 1 repeats our main specification but adjusting the endogenous variable of interest, reducing the migrant pool to those aged 25+ as they are likely to have had the time to attend school. Columns 2 and 3 repeat the specification in Column 1 but including a second endogenous variable for the interaction of immigration with the (demeaned) average years of education or (demeaned) average years of college education, respectively, of those migrants. In all three of these specifications, for instrumentation we exploit the fact that in our initial instrument construction we created quasi-exogenous immigration shocks for each origin country-o $\times$ destination county-d pair in each time period t, $\hat{I}_{o,d}^t = \hat{A}_{o,d}^{t-1} \times \tilde{I}_{o,-r(d)}^t$; each specification utilizes the predicted immigration shocks for each of the the top 20 origin nations as a joint set of instruments for both total immigration and, in the case of Column 2 and Column 3, immigration interacted with average education. For each specification we report the first-stage F-statistic(s), utilizing the F-statistic described in Angrist and Pischke (2009, p. 217-218) in the case of multiple endogenous variables. Standard errors are clustered by state for all specifications and *,**, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

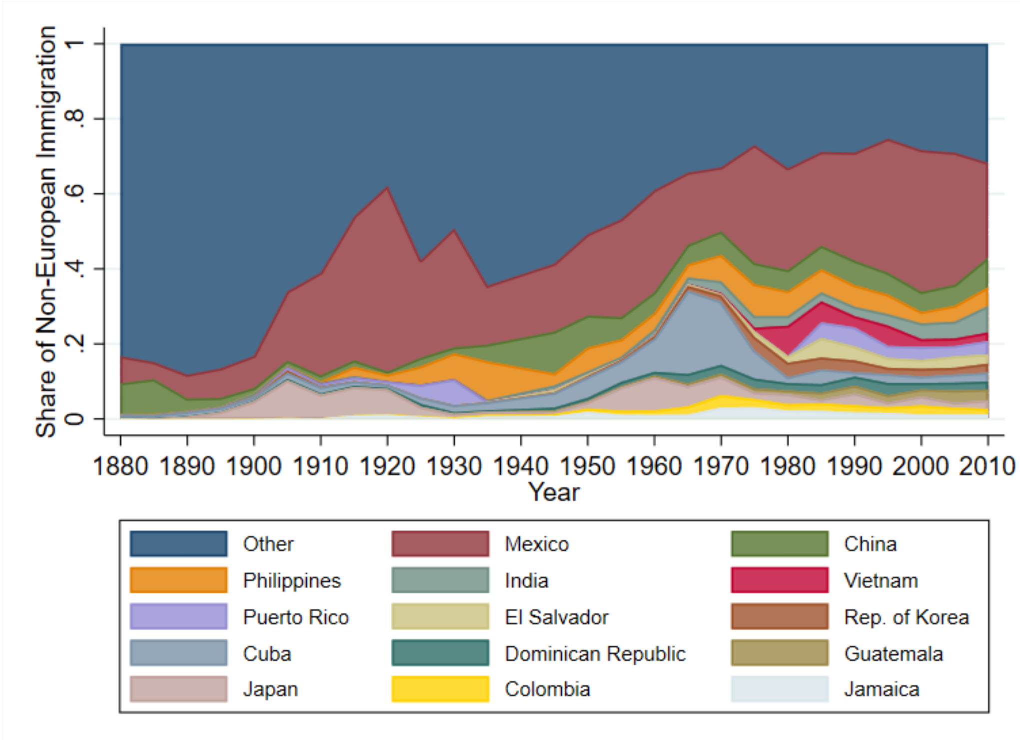

Figure 1: Origins of Non-European Immigrants to the U.S.

Notes: This figure plots the share of non-European immigration into the US from the 14 non-European origin nation (except for Canada which is included in "Other") with the largest cumulative immigration to the US. The figure highlights variation in the push factor, showing that the number of migrants from a given source country o to the US varies by period t.

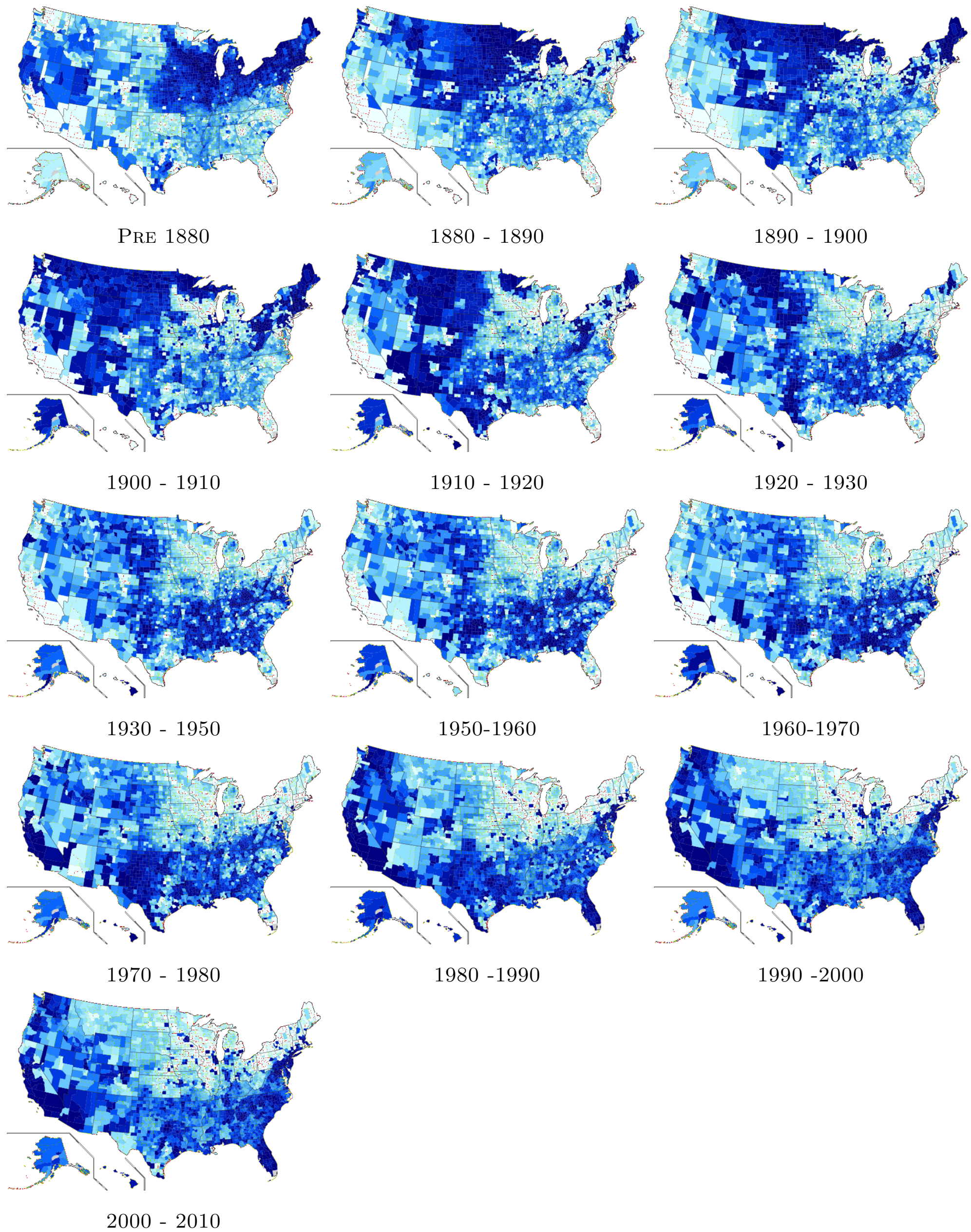

FIGURE 2: DESTINATIONS OF IMMIGRANTS TO THE UNITED STATES, PRE-1880 TO 2000

Notes: This figure maps immigration flows into US counties by 10-year periods (except between 1930 and 1950). We regress the number of immigrants into US county d at time t, I_d^t, on destination county d and year t fixed effects, and calculate the residuals. The map's color coding depicts the residuals' decile across counties and within census periods. Darker colors indicate a higher decile.

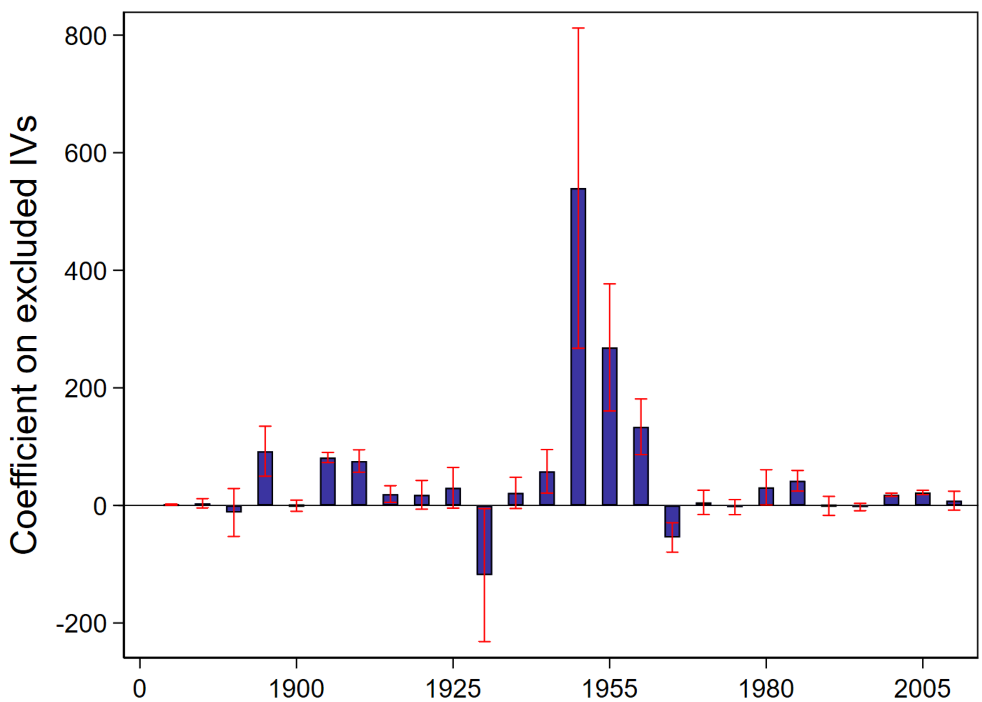

Figure 3: Stage 0: Predicting Ancestry

Notes: Red lines give 95% confidence intervals. Standard errors are clustered at the origin country level. (F-stat 32,645.9)

Notes: This figure displays the coefficients and standard errors in the ancestry prediction regression, equation (3), for estimating 2010 reported ancestry (assuming for presentational purposes only that $b^{\tau}_{r(d)} = b^{\tau} \; \forall r(d)$). The figure shows that we identify variation in current ancestry levels based on push-pull interactions from the full range of time periods in our sample.

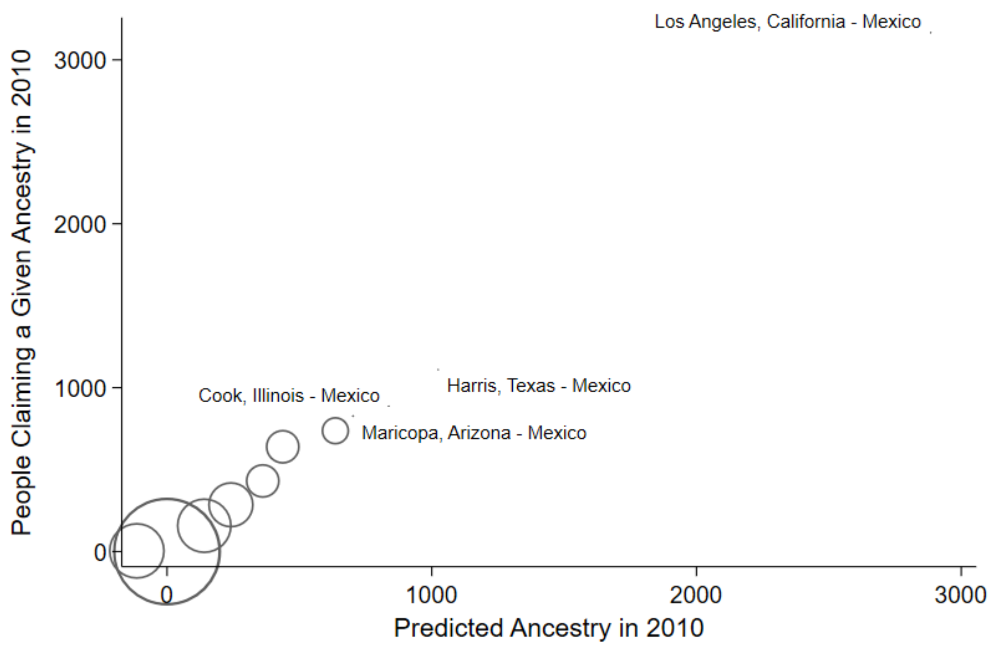

FIGURE 4: STAGE 0: PREDICTING ANCESTRY (2010)

Notes: This figure plots actual ancestry in 2010 against predicted ancestry, as given in equation (3), with the size of each circle indicating the log number of observations in a given bin of predicted ancestry. The labeled counties are those with the highest number of individuals declaring a given ancestry in 2010.

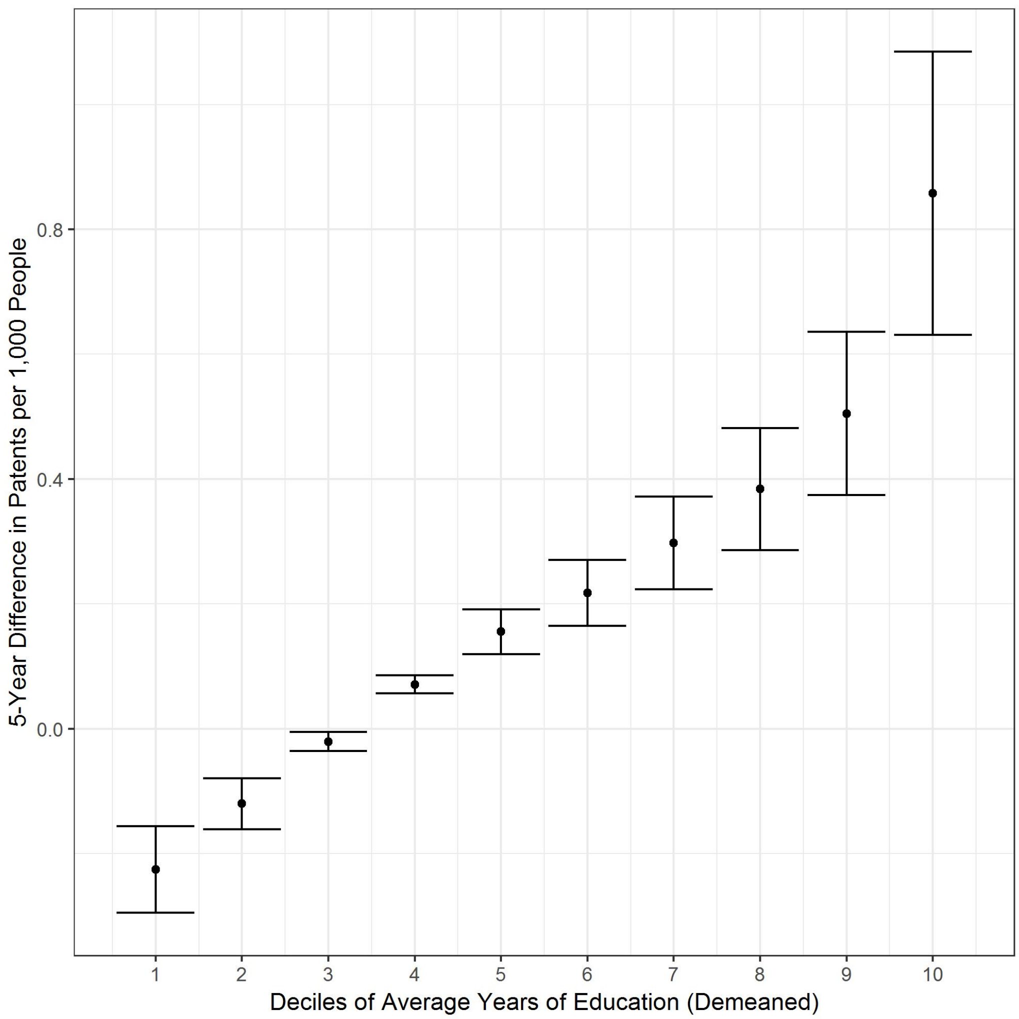

Figure 5: Impact of Immigration and Total Education on Innovation

Notes: This figure plots the impact of immigration on patenting per capita at deciles of schooling implied by the estimates in Column 2 of Table 10.

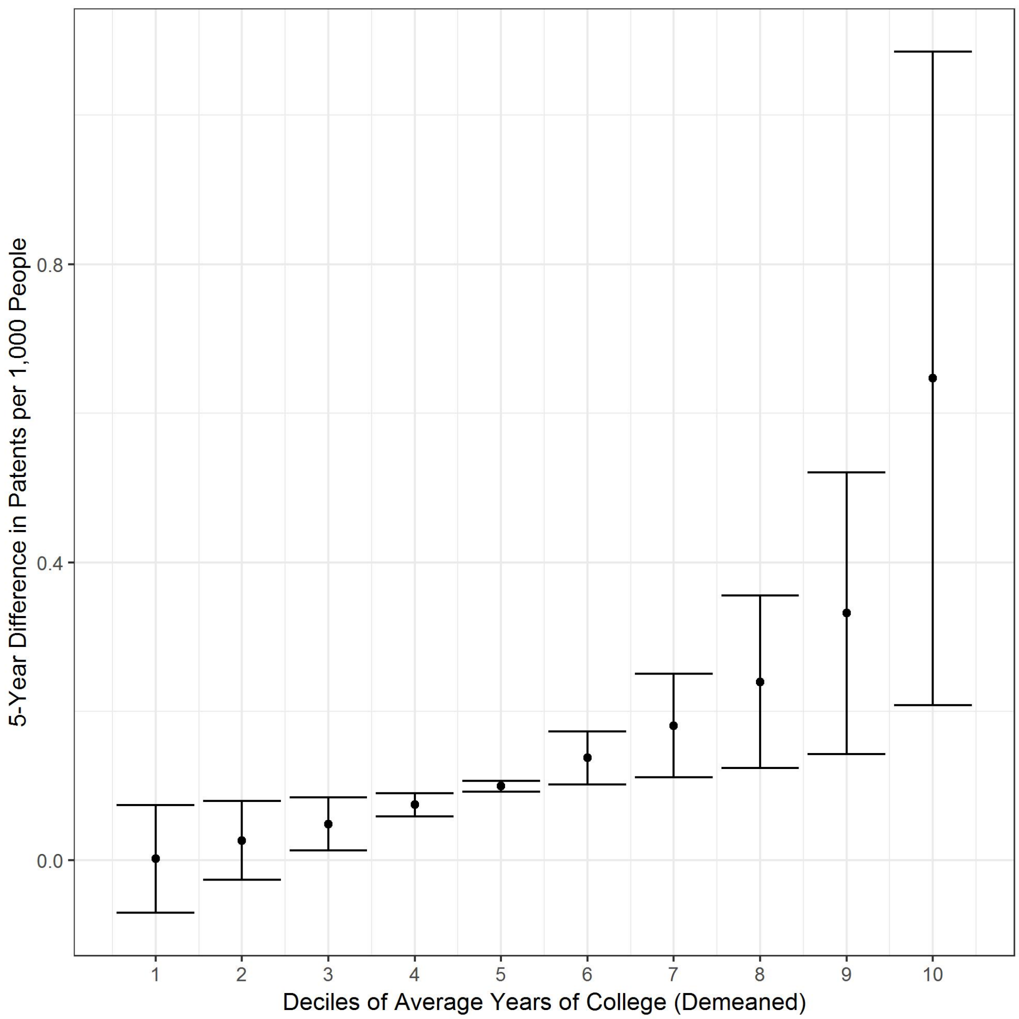

Figure 6: Impact of Immigration and College Education on Innovation

Notes: This figure plots the impact of immigration on patenting per capita at deciles of college education implied by the estimates in Column 3 of Table 10.

Online Appendix

"Immigration, Innovation, and Growth"

Konrad B. Burchardi

Thomas Chaney

Tarek A. Hassan

Lisa Tarquinio

Stephen J. Terry

A Data Appendix

A.1 Details on the construction of migration and ethnicity data

Details calculation of post-1880 flow of immigrants

For each census wave after 1880, we count the number of individuals in each historic US county d who were born in historic country o (as identified by birthplace variable "bpld" in the raw data) that had immigrated to the United States since the last census wave that contains the immigration variable (not always 10 years earlier). Then we transform these data

- from the non-1990 foreign-country ("bpld") level to the 1990 foreign-country level using bpld-to-country transition matrices.
- from the US-county group/puma level to the US-county level using group/puma-to-county transition matrices.
- from the non-1990 US-county level to the 1990 US-county level using county-to-county transition matrices.
- from the post-1990 US-county level to the 1990 US county level. Based on the information from https://www.census.gov/geo/reference/county-changes.html, a new county is either created from part of ONE 1990 county or assigned a new FIPS code after 1990, so we manually change that county's FIPS code to what it was in 1990. A few counties' boundaries have been changed after 1990 but that only involved a tiny change in population, so we ignore these differences.

Details calculation of pre-1880 stock of immigrants

For the year 1880, we calculate for each historic US county d the number of individuals who were born in a historic foreign country o (no matter when they immigrated). We add to those calculations the number of individuals in county d who were born in the United States, but whose parents were born in historic foreign country o. (If the parents were born in different countries, we count the person as half a person from the mother's place of birth, and half a person from the father's place of birth). Then we transform these data

- from the pre-1880 foreign-country ("bpld") level to the 1990 foreign-country level using the pre-1880 country-to-country transition matrix.
- from the pre-1880 US-county level to the 1990 US-county level using the pre-1880 county-to-county transition matrix.

Details calculation of stock of ancestry (1980, 1990, 2000, and 2010)

For the years 1980, 1990, 2000, and 2010, we calculate for each US county group the number of individuals who state as primary ancestry ("ancestr1" variable) some nationality/area. We transform the data

- from the ancestry-answer ("ancestr1") level to the 1990 foreign-country level using ancestry-to-country transition matrices.
- from the US-county group/puma level to the US county-level using group/puma-to-county transition matrices.
- from the non-1990 US-county level to the 1990 US-county level using county-to-county transition matrices.
- from the post-1990 US-county to the 1990 US-county level. Based on the information from `https://www.census.gov/geo/reference/county-changes.html`, a new county is either created from part of ONE 1990 county or assigned a new FIPS code after 1990, so we manually change that county's FIPS code to what it was in 1990. A few counties' boundaries have been changed after 1990 but that only involved a tiny change in population, so we ignore the difference.

Appendix Table 1: Description of each IPUMS wave

Wave	Description
1880	We use the 10% sample with oversamples; the sample is weighted, so we use the provided person weights to get to a representative sample; we use the region identifiers statefip and county.
1900	We use the 5% sample; the sample is weighted, so we use the provided person weights to get to a representative sample; we use the region identifiers statefip and county.
1910	We use the 1% sample; the sample is unweighted; we use the region identifiers statefip and county.
1920	We use the 1% sample; the sample is weighted, so we use the provided person weights to get to a representative sample; we use the region identifiers statefip and county.
1930	We use the 5% sample; the sample is weighted, so we use the provided person weights to get to a representative sample; we use the region identifiers statefip and county.
1970	We use the 1% Form 1 Metro sample; the sample is unweighted; we use the region identifiers statefip and cntygp97 (county group 1970); note that only four states can be completely identified because metropolitan areas that straddle state boundaries are not assigned to states; identifies every metropolitan area of 250,000 or more.
1980	We use the 5% State sample; the sample is unweighted; we use the region identifiers statefip and cntygp98 (county group 1980); the sample identifies all states, larger metropolitan areas, and most counties over 100,000 population.
1990	We use the 5% State sample; the sample is weighted, so we use the provided person weights to get to a representative sample; we use the region identifiers statefip and puma; the sample identifies all states, and within states, most counties or parts of counties with 100,000 or more population.
2000	We use the 5% Census sample; the sample is weighted, so we use the provided person weights to get to a representative sample; we use region identifiers statefip and puma; the sample identifies all states, and within states, most counties or parts of counties with 100,000 or more population.
2010	We use the American Community Service (ACS) 5-Year sample; the sample is weighted, so we use the provided person weights to get to a representative sample; we use region identifiers statefip and puma, which contain at least 100,000 persons; the 2006-2010 data contains all households and persons from the 1% ACS samples for 2006, 2007, 2008, 2009 and 2010, identifiable by year.

Appendix Table 2: Historical birthplace to current country: transition matrices

Panel A: Birthplace		weights $\in (0,1)$	weight $= 1$	weights $= 0$
1880	# of answers	22	258	9
	# of persons	26,301	50,177,184	4,933
	% of persons	0.05%	99.94%	.01%
1900	# of answers	15	131	6
	# of persons	23,345	6,555,140	5,339
	% of persons	0.35%	99.56%	.08%
1910	# of answers	20	99	4
	# of persons	31,072	5,613,136	3,105
	% of persons	0.55%	99.39%	.05%
1920	# of answers	13	174	7
	# of persons	36,070	3,905,455	12,559
	% of persons	0.91%	98.77%	.32%
1930	# of answers	25	194	9
	# of persons	35,930	3,086,341	61,462
	% of persons	1.13%	96.94%	1.93%
1970	# of answers	12	77	3
	# of persons	318,800	6,323,100	230,800
	% of persons	4.64%	92.00%	3.36%
1980	# of answers	32	222	7
	# of persons	491,760	4,774,820	313,300
	% of persons	8.81%	85.57%	5.61%
1990	# of answers	24	209	7
	# of persons	721,595	8,532,585	484,433
	% of persons	7.41%	87.62%	4.97%
2000	# of answers	11	136	0
	# of persons	1,122,532	13,144,632	0
	% of persons	7.87%	92.13%	0%
2010	# of answers	14	137	1
	# of persons	1,302,255	11,131,046	17,148
	% of persons	10.46%	89.40%	.14%
2010*	# of answers	14	188	1
	# of persons	3,512,123	300,415,680	37,469
	% of persons	1.16%	98.83%	.01%
Panel B: Ancestry		**weights $\in (0,1)$**	**weight $= 1$**	**weights $= 0$**
1980	# of answers	29	227	143
	# of persons	924,400	198,525,616	27,412,380
	% of persons	0.41%	87.51%	12.08%
1990	# of answers	29	239	9
	# of persons	2,941,941	217,720,512	27,445,182
	% of persons	1.19%	87.75%	11.06%
2000	# of answers	17	137	22
	# of persons	6,000,639	191,300,704	84,120,558
	% of persons	2.13%	67.98%	29.9%
2010	# of answers	19	142	30
	# of persons	8,454,279	229,211,968	66,299,030
	% of persons	2.78%	75.41%	21.81%

The table reports statistics on the transition of data from the 'answer' level to 1990 country level. For each survey wave, and each question – birthplace in Panel A and primary ancestry in Panel B – the table reports the number of answers that can be directly linked to a 1990 country (weight = 1), that are assigned to several 1990 countries using population weights (weights $\in (0,1)$) and that cannot be linked to any modern country with sufficient certainty (weights = 0). The table also reports the number of respondents (scaled from the original data using the person weights provided) in each category. Answers with weights zero essentially consists of "Not Reported" (e.g. 23, 24, 54 and 30 million respondents for the 1980, 1990, 2000 and 2010 ancestry data, respectively) and "African-American" (e.g. 26, 22 and 25 million respondents for the 1990, 2000 and 2010 ancestry data, respectively). The remainders are mostly cases such as "African", "Uncodable", "Bohemian", "Nuevo Mexicano", "Other", etc. In Panel A, all years except 1880 consist of the number of persons that report birthplace since the last Census wave. For the 2010 Census wave the additional entry (denoted by a *) reports the respective numbers for all respondents in that wave.

Appendix Table 3: Historical state-county unit to 1990 state-county unit: transition matrices

Census wave		weights $\in (0,1)$	weight $= 1$	weights $= 0$
1880	# of counties	658	1854	1
	% of persons (birthplace data)	21.54%	78.45%	.01%
1900	# of counties	2211	7	4
	% of persons (birthplace data)	99.09%	0.87%	.05%
1910	# of counties	1517	5	1
	% of persons (birthplace data)	99.00%	0.94%	.05%
1920	# of counties	1355	7	0
	% of persons (birthplace data)	90.80%	9.20%	0%
1930	# of counties	1801	6	0
	% of persons (birthplace data)	90.61%	9.39%	0%
1970	# of countygroups	310	98	0
	% of persons (birthplace data)	34.07%	65.93%	0%
1980	# of countygroups	580	573	0
	% of persons (birthplace data)	17.96%	82.04%	0%
	% of persons (ancestry data)	40.02%	59.98%	0%
1990	# of PUMAs	541	1185	0
	% of persons (birthplace data)	8.97%	91.03%	0%
	% of persons (ancestry data)	32.15%	67.85%	0%
2000	# of PUMAs	620	1451	0
	% of persons (birthplace data)	10.66%	89.34%	0%
	% of persons (ancestry data)	30.36%	69.64%	0%
2010	# of PUMAs	619	1449	1
	% of persons (birthplace data)	12.31%	87.65%	.03%
	% of persons (ancestry data)	30.13%	69.81%	.05%

The table reports statistics on the transition of data from the 'historical spatial area' level to 1990 US county level. For each Census wave the table reports the number of contemporaneous spatial areas that are a subset of a 1990 US county (weight = 1) and the number of contemporaneous spatial areas whose data is transitioned to 1990 US county level using non-degenerate weights (weights $\in (0,1)$). For Census waves 1880 to 1930 the share of their contemporaneous county spatial area in each 1990 US county area is used as weight. For waves 1970 to 2010 there are two steps: In step 1 the share of their contemporaneous countygroup (waves 1970 and 1980) or PUMA (waves 1990 to 2010) population in the contemporaneous county population are used as weights; in step 2 the share of their contemporaneous county spatial area in each 1990 US county area is used as weight. The two-step procedure is necessary because the 1970 to 2010 Census waves do not have a county-level identifier (to protect the privacy of the respondents). The table also reports the share of respondents affected by this transition in the birthplace and ancestry data, respectively.

B Growth, Population Growth, Innovation, & Dynamism

In this appendix we sketch out a deliberately simple theoretical mechanism linking innovation, income growth, dynamism, and population growth. We present the minimum ingredients needed from a combination of the "semi-endogenous growth" model outlined in Jones (1995) and the micro-level distribution of creative destruction from Schumpeterian growth models (Aghion and Howitt, 1992; Grossman and Helpman, 1991; Klette and Kortum, 2004). We show that in such a model the long-run balanced growth path per capita growth rate of the economy must be proportional to the growth rate of labor input in the economy and that the economy-wide growth rate links positively to the rates of creative destruction and innovation at the micro level. These two outcomes concisely justify our empirical analysis linking population dynamics to measures of scaled innovation, dynamism rates, and income growth, abstracting from cross-economy spillovers and heterogeneity in labor input, both of which we nevertheless explore empirically.

B.1 Environment

Final Goods Production We examine a closed local economy in continuous time t. Final output Y_t is produced according to the technology

$$\log Y_t = \int \log y_{jt} dj$$

utilizing a unit mass of intermediate varieties j.

Intermediate Goods Production Intermediate goods are each produced with a symmetric technology combining production labor l^P_{jt} and variety-specific quality q_{jt}, with $y_{jt} = q_{jt} l^P_{jt}$. Incumbent intermediate goods firms f produce portfolios of intermediate varieties j for which they operate the current leading-edge quality level q_{jt}. Let $\log Q_t = \int \log q_{jt} dj$ be the average quality level in the economy.

Innovation For an individual variety, innovation is embodied in instantaneous increase in the quality level q_{jt} in that good's production, i.e., a switch from q_{jt} to $q_{jt+\Delta} = \lambda q_{jt}$, where $\lambda > 1$ is a quality ladder or innovation step size. Incumbent firms f may innovate by hiring labor for innovation in the amount s^I_{ft} to guarantee an innovation arrival rate p^I_{ft} satisfying

$$p^I_{ft} \propto {s^I_{ft}}^\gamma Q_t^{-\alpha},$$

where $\alpha, \gamma > 0$. A mass of potential entrants each hires labor for innovation s^E_t to guarantee an innovation arrival rate p^E_t satisfying

$$p^E_t \propto s^{E\gamma}_t Q_t^{-\alpha}.$$

In both of the innovation technologies, innovation arrival probabilities depend positively on innovation input – labor – but negatively on the current average quality level in the economy Q_t. Solving harder problems to improve upon a higher existing average quality level requires more input. When an innovation occurs, for either an entrant or incumbent, they become the leading-edge incumbent producer of a random variety.

Labor Input The exogenous instantaneous growth rate of labor input or the population of the economy L_t is n, and total labor input in any period must equal the sum of the total amounts of labor used for production, incumbent innovation, and entrant innovation.

$$L_t = L^P_t + S^I_t + S^E_t$$

B.2 Balanced Growth

A range of straightforward and standard additional machinery needed for description of a decentralized equilibrium along a stationary balanced growth path – along the lines of the equilibria described in Klette and Kortum (2004) or Grossman and Helpman (1991) – could be added to the framework already outlined above. But we do not need additional elements for our desired implications. Instead, we simply note that in standard decentralizations output per capita is proportional to the average quality level Q_t. We also note that along any stationary balanced growth path in this economy by definition there must be constant output growth rates, constant quality growth rates, constant ratios of production labor and innovation labor to total labor input, and a stationary distribution of outcomes at the firm and variety levels.

But then note that constant quality growth rates and constant innovation rates for incumbents and entrants - given the innovation technologies - imply that

$$Q^\alpha_t \propto S^{I\gamma}_t \propto S^{E\gamma}_t \propto L^\gamma_t \rightarrow \alpha g_Q = \gamma n \rightarrow g_Q = \frac{\gamma}{\alpha} n.$$

In other words, average quality growth, which is equal to per capita growth in this economy, must be positively proportional to the population growth rate n. This is our first desired result, echoing Jones (1995). Then, given the definition of average quality Q_t, the implication of a constant growth rate $g_Q = \frac{\partial \log Q_t}{\partial t}$ is that

$$g_Q = p \log \lambda,$$

where $p = p^I + p^E$ is the sum of the constant incumbent and entrant innovation rates and λ is the quality ladder step size described above. But note that

$$p = \mathbb{P}(\text{Innovation}) = \mathbb{P}(\text{Displacement})$$

in this Schumpeterian economy. So we obtain that

$$\mathbb{P}(\text{Innovation}) = \mathbb{P}(\text{Displacement}) = \frac{g_Q}{\log \lambda} = \frac{\gamma}{\alpha \log \lambda} n,$$

i.e., the rate of creative destruction and the innovation rate are positively proportional to population growth. This is our second result, following directly from the logic of creative destruction-based growth models.

B.3 Implications

Along a balanced growth path, in models with the ingredients outlined above, we must have the following implications.

- Per-capita output and income growth rates positively link to population growth rates.
- Innovation rates positively link to population growth rates.
- Creative destruction or displacement rates positively link to population growth rates.